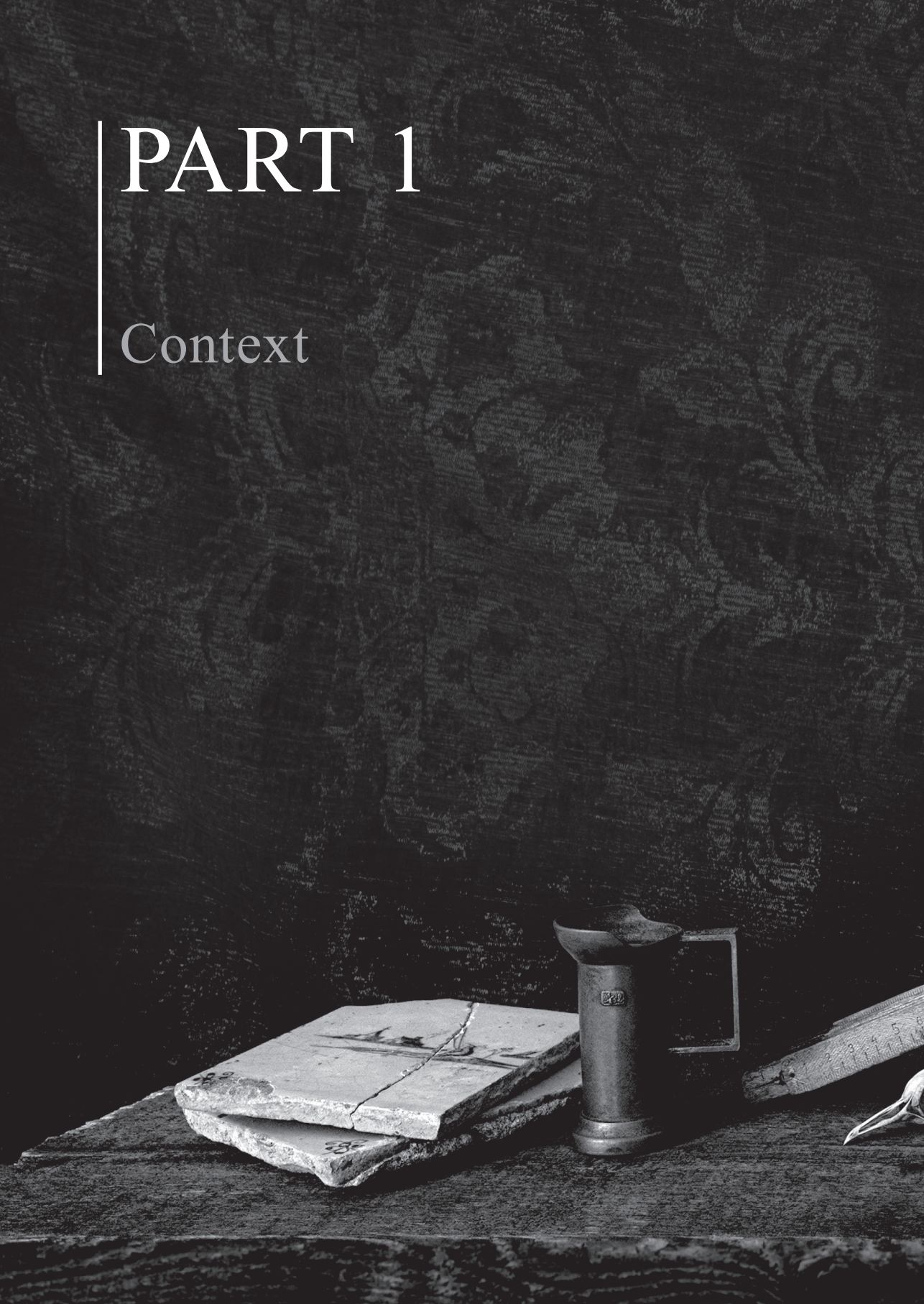
Computer says no: understanding digital authority

Weggemans, D.J.

PART 1

Context



1 INTRODUCTION

A Belgian woman wanted to visit her friend in Brussels. A one-hour drive, but somehow she ended up 900 miles off-route. In a newspaper article the woman described how she drove for hours and noticed how the language of the road signs changed from Dutch to French and then to German - Köln, Frankfurt, München (De Standaard, 2013). She did not question her navigation system and kept driving. During her trip she stopped for gas twice and slept in her car at the side of the road. Eventually, she found herself in Zagreb, Croatia, and realized she was completely in the wrong place.[1]

The woman definitely is not alone in following her GPS into disaster. In Colorado nearly a hundred drivers got stuck in an empty, muddy field after Google Maps suggested them a detour route (CNN, 2019). In Alaska, several motorists ignored many signs, lights, and pavement markings to follow their navigation app onto the runway of an airport (BBC, 2013a). In Brazil, there are various cases of people who were shot after their GPS sent them into violent favelas instead of the beach (The Washington Post, 2015). And from all over the globe there are countless stories of people driving their vehicles straight into public parks, houses, oceans, rivers, sand pits and off cliffs' edges because of the instructions of a machine.

In a matter of decades, our lives have been flooded with incredible, communicating, digital technologies. Computers that can inform, advise, recommend, tell, instruct, and order. We find them in a myriad of decision-making contexts ranging from our homes to the office and from the road to leisure. And most of the time, relying on them makes perfect sense. After all, they are designed to address everyday challenges. Unfortunately, no technology is flawless. There can always be bugs, bad algorithms, biased data, and input errors that can lead us astray. But while most of us are aware of this; we remain all too eager to do what we are digitally told. And this goes well beyond the moments we are in our cars, guided by a navigation system. The phenomenon of people's unquestioning acceptance of a technology's output has become a frequent subject of media reports, as well as a popular theme for satire. For instance, in the comedy television show *Little Britain* there is the character of a woman sitting at a desk. When her clients describe what they want, she starts typing, resulting in the same response every time: "computer says no". The phrase has become iconic for describing those situations in which people are led by whatever a computer tells them.

1.1 DIGITAL SECURITY AND BEYOND

Security increasingly runs on algorithms and data. Technological developments and new applications allow for many decisions in the field of security to revolve around the digital collection, analysis, and use of large amounts of data. Large and linked real-time datasets informing decision-making processes have become commonplace. Infractions are digitally detected, and fines are (semi-)automatically processed on a large scale – such as traffic violations and tax or benefits fraud. And digital technologies now form an integral part of current efforts to predict and pre-empt all sorts of risks. Police forces, for example, embrace predictive algorithms that analyse historical and criminal data to estimate who will commit future criminal offences or where they might occur (Meijer, Lorenz & Wessels, 2021: 839).[2] Courts and other organisations in the criminal justice system increasingly rely on algorithms and data to predict the future dangerousness of individuals - there are hundreds of risk assessment tools available that are widely used to inform pretrial decisions, criminal sentencing, parole, and post release monitoring (Peeters & Schuilenburg, 2018: 272; Hamilton, 2015; Stevensen, 2018). Furtermore, digital technologies are also used to identify those at risk, for instance, in sectors such as child welfare (Keymolen & Broeders, 2013; Eubanks, 2018; Hurley, 2018; Redden, Denick & Warne, 2020).

The term *digital security governance* captures this approach to working on security issues. Broadly speaking, it is the use of a range of digital technologies for the collection, processing, storage, sharing and application of information for security purposes.[3] The term not just highlights how digitisation has allowed for the automation of (parts of) security governance. Digital security governance also covers the many digital applications that provide advice, suggestions, or information to a human decision-maker. The increased role of digital technologies has significantly changed the nature of decision-making in many security contexts. Many decisions are now made in the presence of digital technologies that provide human decision-makers with a constant flow of input on what to do or think. This entails that digital technologies have become directly involved in decisions that substantially impact human lives.

People's blind reliance on erroneous digital output can thus have potentially far-reaching consequences. While a GPS might send us in the wrong direction, or a sports game can be unfairly influenced by a referee's deference to a mistaken call of the Hawk-eye technology[4], the impact of digital technologies involved in more critical affairs can be more problematic, even life-threatening. We find examples across sectors. In medicine, for instance, recommendations of a computer system caused highly trained professionals to misdiagnose patients or administer lethal doses of medication (Wachter, 2015). In the financial sector, overreliance on mistaken algorithms has led to invalidly foreclosed homes (Reuters, 2018). In immigration, legal citizens have been apprehended because of wrong tags in key digital systems (Kohli, Markowitz & Chavez, 2011). And in the criminal justice system, dedication to the words of digital technologies resulted in wrongful arrests (Jopson, 2010; Jacobsen, 2011;

Joh, 2016) and releases (see chapter 4). There are even reports of police shootings based on inaccurate digital information about vehicle theft (Van den Hoven, 1998:98).

With cutting edge technology being constantly sought by governments and businesses for security and other important issues, people's willingness to listen to a computer becomes more intriguing as well as more disturbing. How far can we go? And why?

1.2 DECISION-MAKING IN A WORLD OF COMPLEXITY

This book is about making important decisions in a digitising world. Most of these decisions are characterised by complexity and uncertainty.[5] Human decision-makers are confronted with the unforeseen actions of others, different possible futures, and the risk of *information overload* - having to handle more information than can be processed. This prevents people from making perfectly rational decisions. There is simply too much going on, and human cognitive capacities do not allow for systematically thinking through the many consequences of each choice.[6]

We have different strategies to deal with this complexity. Some are *internal*, cognitive, strategies to simplify decision-making. For example, people often look for the path of least cognitive resistance – just like they do with physical tasks. That is, they seek to think and solve problems in ways that are simple and do not take much cognitive effort (Mosier & Skitka, 2009: 203; Fiske & Taylor, 1994).[7] Social psychologists have highlighted that for many decisions – both important and inconsequential – we therefore use *heuristics* (Simon, 1972; Kahneman & Tversky, 1973; Gigerenzer et al, 2011). These are simple decision-making rules, or "rules of thumb", used to form judgements or draw conclusions quickly and easily. While these are not always right or the most accurate, they do help to find answers and solutions that are accurate enough and work most of the time (Kahneman & Tversky, 1973: 237).

We have also always used strategies to cope with life's complexities that are more *external*, by creating technological artifacts. Things like calendars, notebooks, calculators, checklists, maps, and clocks have been effective to overcome - or at least effectively deal with - the limitations of our minds. And in today's world, the digital technologies that have permeated many human decision-making contexts are approached in a similar way. Because they take over certain tasks, make relevant information accessible when needed or simply tell us what to do and think, they are often viewed to support and augment human decision-making processes (Van den Hoven, 1998; Zack, 2007; Woods, 2009; Noyes, Cook & Masakowski, 2016).

But these same digital technologies also bring forth new complexities. Research at the turn of the century already indicated that digital technologies could become a highly salient source of information for decision-makers (Parasuram & Riley, 1997; Skitka, Mosier & Burdick, 2000; Skitka, Mosier, Burdick & Rosenblatt, 2000). Digital cues draw and

focus our attention on many occasions and even become a "heuristic replacement for vigilant information seeking and processing" (Mosier & Skitka, 2009: 203).[8] Without much cognitive effort, the additional suggestions of a computer are then turned into a user's final decision (Citron, 2008: 1272). However, as mentioned, computer systems are still not the infallible super-brains we have been hoping for. While they allow us to do many new things, they are also quite dumb. They are machines that strongly rely on the data we feed them and the instructions we provide them with. And when these data are flawed or the algorithms are incorrect, the technology will tell us the wrong thing. The paradoxical result is that digital support for human-decision making does not always help people make better decisions but may also lead them to do remarkably stupid things. Even when it makes no sense at all, we still willingly follow the machine.

1.3 BEYOND TRUST

Much of this has already been studied through the lens of human trust in technological artifacts. Trust, too, is viewed as part of our repertoire for coping with the uncertainties inherent in human life (Keymolen, 2016: 12). Trust involves the willingness of an actor (the trustor) to be *vulnerable* to the actions of another actor (the trustee) on the basis of *expectations* that his or her actions or intentions will be positive or favorable. Trust does not hinge on the trustor's ability to monitor or control this other actor. Ultimately, it is a leap of faith – we can never perfectly predict or be absolutely certain that someone else will perform the particular action important to us (Möllering, 2006: 105-121; Keymolen, 2016: 15-18; Botsman, 2017: 19-20). But, despite that trusting is a risky business as it comes with an *irreducible* social vulnerability and the possibility of getting hurt, it does allow people to act. Without it, people would be paralysed by inaction as they would be constantly confronted with the incomprehensible complexity of considering all possibilities out there for deciding what to do. To reduce the experience of risks and uncertainty we therefore trust in people *and* technologies all the time (Lankton, McKnight & Tripp, 2015).

People can place a great deal of trust in seemingly intelligent technologies and even make their lives depend on them – think of autopilots or surgical robots. But what about those situations in which trust is unclear or only forms part of the explanation for people's compliance to technologies? Take, for example, the moments in which people willingly permit their behavior to be determined by a technology. While this often occurs in conjunction with trust, this is not necessarily so. Throughout this book there are examples of situations in which people felt they could simply not ignore the directives of a computer or gave up their own judgement (and sense of responsibility) to do what they were digitally told. Even when they did not trust the output and believe the technology is wrong. This is also illustrated by a group of cycling tourists who ended up in an Amsterdam highway tunnel. At the end of the tunnel a coincidentally passing television reporter interviewed the tourists:

> "I don't know what happened. I followed the navigation" the cyclist starts.
> "It showed that there was a bicycle lane in the tunnel. (..) I saw all kinds of red signs and remember thinking: you are not allowed to go here. But look at the navigation! I don't want to come across as an idiot, but it really says: 'enter the tunnel!'. (..) We could have died. Imagine that a car would have come our way! This is dangerous, Apple! [expletive, expletive]. It was scary. People now probably think that we are idiots, but it was Apple sending us this way!".
> Then the police arrived. "I blame my navigation" he begins.
> "But you also have brains, right?" the police officer replies while pointing at her head.
> The tourist then turns to the camera: "They blame us. But do you think that is fair?"[9]

These compliant, or deferent, human responses are worthy of deeper investigation and shift our focus to the influence of the technology itself. As it appears, digital technologies can outgrow their role as trusted digital assistants and compellingly influence us in what we think and do. This, in turn, presents us with a curious dilemma because were not technologies just means for human beings to better carry out their tasks?

1.4 THE INFLUENCE OF THINGS

Technologies are commonly viewed as tools or instruments (Feenberg, 1991: 6). In this view, a technology is defined by how it can be utilised for specific ends. The technology itself remains a neutral object. It is neither good or bad, rather it is what we *do* with the technology that makes it good or bad (or a bit of both). The technology is merely an instrument that we can use in the way we want. It is a means to an end (Van den Berg, 2009: 28). For instance, digital technologies are often considered as an important weapon against all sorts of risks – used by people to prevent and detect threats and take necessary measures to increase security.

This instrumentalist view in which we are the ones doing things to technologies, however, overlooks the fact that they do things to us as well (Verbeek, 2005). It underestimates technological artifacts and the various ways they influence us. Philosophers of technology have already written extensively on the role of technology in shaping our experiences and interpretations of the world (Heidegger, 1927; Merleau-Ponty, 1962; Ihde, 1990; Verbeek, 2005). For example, technologies can amplify aspects of reality while reducing others – think of looking at people through full body scanners at airports which reveals objects on or inside a person's body for security purposes, while at the same time distorting much of what is visible to the naked human eye (see also Verbeek, 2011:9).[10] Similarly, there have been detailed investigations into how technological artifacts influence our actions and

decisions. Latour (1992) and Akrich (1992) introduced the concept of "script" to describe the influence of technologies on human behavior.[11] Like movie or theater scripts that guide actors, technologies can prescribe their users how to act. A plastic coffee cup, for instance, has the script "throw me away after use", while a porcelain cup *suggests* to be cleaned and used again (c.f. Verbeek, 2006a: 362). And sometimes such technological scripts are more *coercive*: speedbumps, one of Bruno Latour's (1992: 244) well-known examples, pressure drivers to slow down for they will otherwise destroy the car's suspension.

The development of our digital age has generated new modes of technological influence. Computers are not just important channels for interaction with other people but can also be a source of interaction in themselves. That is, we no longer only interact *through* technologies, but also, for the first time in history, *with* them (Ihde, 1990; Fogg, 2003). And in these new human-technology configurations, computers influence people in unprecedented ways. For instance, in *Persuasive Technology* (2003) B.J. Fogg has introduced the concept of *persuasion* – from the classical discipline of rhetoric - to describe how technological artifacts, like discursive arguments, can motivate people into changing their behavior and attitudes (c.f. Dorrestijn & Verbeek, 2013: 46). A speedometer on the side of the road, for example, tries to persuade drivers to stay within limits by displaying how fast they are going – sometimes supplemented with a happy or sad emoticon (Fogg, 2003: 42). Others have highlighted how technologies have also "learnt" to *deceive* (Castelfranchi, 2000). This form of influence is not based on arguments, but rather on misleading information. We all know the windows popping up on our screen that falsely warn us against a detected virus and that we should "click here!" to solve the problem, or the websites that try to trick us into paying an exorbitant price for a concert's last tickets (while plenty are still available elsewhere).

The influence of technologies, of all kinds, has become well-trodden territory. Various approaches and concepts have been developed that help us understand technologies as more than simply instruments of human will. They have helped to shift focus away from the functional characteristics of technologies and direct our attention to other forms of interaction and new human-technology relationships (see also chapter 4). But how about those situations in which we seemingly respond to digital technologies as the Master of Human will? How can we understand the moments when there is no direct technological persuasion, coercion, or deception but some other compelling influence that makes people listen? The answers to these questions, may be found through the lens of authority.

1.5 ON THE CONCEPT OF AUTHORITY

Authority has been described as "extraordinarily complex, hopelessly elusive, and almost as badly misconstrued in most scholarly discussions as it is in popular parlance" (Lincoln, 1994: 1). The concept has many different associations and meanings within various disciplines, including political theory, public administration, sociology, and psychology. And in our daily lives we tend to apply the term rather loosely and as a synonym for power – for example when we say, "someone has the authority to do something". Its prolific and ambiguous use, in a range of contexts, has worked to mystify the meaning of authority.

As a general starting point, I take authority to apply to those situations where there is conformity or compliance without the use of force or a lot of discussion. When there is authority, people may accept a decision or obey an order simply because of who gave it or made it (Peters, 1958: 218). For the analysis of social situations, the term authority, thus, emphasises a special kind of influence that regulates human behavior through particular utterances in contrast to other modes of influence (ibid.). As Hannah Arendt once wrote, "if authority is to be defined at all, then, it must be in contradistinction to both coercion by force and persuasion through argument" (Arendt, 1961: 93). Apparently, authority can make us do or think things without being forced and without being rationally convinced.

This implies that authority always involves some sort of acknowledgement on the part of its subjects that the other should be listened to. Therefore, it is not a quality or characteristic that an actor simply has but is essentially relational. Authority is grounded in a recognised relational asymmetry between at least two actors that ensures the confidence, trust and respect of those subject to it – or at least their willingness to behave as if this were so (Lincoln, 1994:4). While often treated with a sense of suspicion, authority rests with this idea of voluntary recognition and is also always open to contestation and change.

Once people recognise authority, they tend to not actively deliberate over the form or content of what is communicated. In fact, under the sway of authority our own judgements do not seem to matter much for doing or thinking what we are told. Rather, our compliance with authorities often takes place with little or no conscious evaluation. And much of this can be traced back to earlier life experiences where we found out about the benefits of following authorities. Psychologist Robert Cialdini describes that in our childhood we learn that listening to our parents and teachers offers genuine advantages because of their greater knowledge, but also because they controlled our rewards and punishments (Cialdini, 2007: 218). Later in life, we encounter many others – from judges and doctors to military officials and coaches – who we learn to associate with the same "superior access to information and power" (ibid.). This way, Cialdini concludes, "information from a recognised authority can provide us a valuable shortcut for deciding how to act in a situation" (ibid.).

While reliance on authority allows people to do many things they could not have done otherwise, it can also have adverse outcomes. Consider the case of imposters who can effectively wield authority by presenting themselves as authority figures for wrongful

purposes, or when legitimate authorities start making illegitimate proposals. This was also demonstrated by the famous and controversial experiments of Stanley Milgram (1963). Over a twelve-year period, he conducted his *obedience studies* with more than a thousand individuals.[12] Much of this type of research cannot be conducted today – it would not be approved as it does not meet modern ethical standards. Milgram told his subjects that he was researching a method in which learning was stimulated by physical punishment. In his lab, an "experimenter" instructed each subject to deliver increasingly severe shocks to an unseen person in an adjacent room every time that person provided wrong answers. There were screams of pain in apparent response to the shocks, but nobody was being hurt – the victims were actors and did not receive any electric shocks. Yet it explicitly illustrated, the "extreme willingness of adults to go to almost any length on the command of authority" (Milgram, 1974: 5).[13] An outcome that was shocking in itself.

Still, people need authorities for many things and will search for them in many situations. We simply cannot do without authority as there will always be situations in which we do not know what to do or how to do things. Deference to authority is one typical strategy to deal with the challenges of everyday life as described in paragraph 1.2. Throughout our lives we encounter many different authorities whose words we will trust, respect, and listen to. But although we may sometimes agonise over who are and are not rightly sought after as authorities, for a long time, there has been the consensus that authority is a relationship between people.

1.6 AUTHORITY OF TECHNOLOGY?

The concept of authority has remained largely peripherical to the analysis of technology. Our lives are permeated with authorities and the authority relationship is as old as human civilisation itself, but the term is seldom used to refer to technologies. In the minds of many, authorities are human beings such as police officers, teachers, parents, or experts on some subject matter – not things.

In philosophy, the authority of technology has often been interpreted as something that facilitates the *construction* of human (political) authority and power. We find a good example in the work of Langdon Winner (1986). Winner famously argued that many technological artifacts are inherently political as their creation and operation require (or are "strongly compatible with") particular social structures (Winner, 1986: 32). He mentions Plato's illustration of a ship on the high seas, where Plato argues that "no reasonable person believes that ships can be run democratically" as their operation requires centralised rule and coordination of many individuals (ibid, 31). In contrast to a one-person boat, Plato took it a practical necessity that large ships have one captain and an unquestionably obedient crew. Winner also refers to Friedrich Engels who pointed out that complicated technical systems, such as big factories, can incentivise more centralised control. When systems get more

complex, according to Engels, "laborers must learn to accept a rigid discipline" and "central control by knowledgeable people acting at the top of a rigid social hierarchy would seem increasingly prudent" (Winner, 1986: 30-31). In a similar vein, Jerry Mander argued that "if you accept nuclear power plants, you also accept a techno-scientific industrial military elite. Without these people in charge, you could not have nuclear power" (Winner, 1986: 32). In this interpretation, technologies have authority in that they structure social hierarchies - their use comes with relational asymmetries between people and allows some to claim authority.

There are also instances in which technologies have been taken as a means for the *exercise* of power by so-called authorities. That is, those in positions of authority may rely on technological artifacts to get things done. Winner's discussion of the low-hanging overpasses on Long Island (New York) parkways that led to the beach illustrates this well. He describes how Robert Moses, an American urban planner and public official, deliberately built these overpasses so low that buses could not use the parkways in the area (Winner, 1986: 22-3). One consequence was the limitation of access for racial minorities and low-income groups to the beach as they normally used public transit and could not afford cars of their own. This example highlights how the design or use of a specific technical system can become a way of settling an issue (ibid: 22).

While these interpretations touch upon (parts of) the issue of authority, they do not directly consider the authority of things themselves. Instead, technologies are either taken as an instrumental intermediary between a human authority and their subjects (i.e., the Long Island overpasses) or as a factor that potentially gives rise to authority relationships between people. The possibility of things *as* authorities, compellingly influencing the thoughts and behaviors of people, is generally overlooked, or dismissed. Of course, sometimes someone talks about the authority of the law or a book, but this too tends to be typically considered as the authority of people (or God) "objectified or embodied in human products" as De George (1985: 16) indicates. (See also McLaughin, 2007: 44).[14] For most, authority remains an exclusively human affair.

And for a long time, there was no reason to suggest otherwise. Surely, as emphasised above, technologies have always been influential and capable of influencing our perceptions and actions. However, technologies did not explicitly give us commands or advice to be followed. It did not, therefore, make much sense to consider the authority of technological artifacts such as hammers or brooms (see paragraph 4.3). But with technological digitisation, this fundamentally changed. Digital technologies are now hyperconnected, increasingly capable and can access a world of data (see paragraph 2.1). This has, among other things, resulted in more and more interactions between people and technologies at any time and any place.

In the *Second Self: Computers and the human spirit* (2005), Sherry Turkle has described digital technologies as evocative objects that incite us to rethink common beliefs about what artifacts are and what it means to be human. These artifacts challenge a deeply held ontological assumption about the difference between the material and the social.

That is, as Suchman elaborates elsewhere, the typical distinction that is made between the things that one uses and the things one communicates with (Suchman, 2007: 33-4). The advent of digital technologies has made that people now have come to *interact* with machines. For the first time, things now issue communications to be believed, obeyed, or followed. And, for the first time, we now see all sorts of unquestioning, docile, and accepting human responses to the communications of technical *things*. Responses that are typically associated with authority.

In recent years, this process of digitisation has incited some to talk about the authority of modern-day technologies. Scholars have, for example, used terms as algorithmic authority (Lustig & Nardi, 2015; Lupton & Jutel, 2015; Lupton, 2016; Lustig et al, 2016; McQuillan, 2016; Jongepier; 2020) artificial authority (Van den Hoven, 1998; French & Lindsay, 2002), robotic authority (Geiskokovitch, et al., 2016; Cormier et al., 2013) and machine authority (Schwarz, 2017; Haring, et al., 2019). These terms were coined in different contexts, focus on particular technologies, and in most discussions, their usage tends to take the form of passing references. However, what they all reveal is the emerging shift in thinking about the nature of authority by explicitly linking it to technological artifacts.

To understand our contemporary world, Science and Technology scholars often resort to borrowing concepts from other fields and, it is worth investigating if authority is a useful addition to this conceptual toolbox. But while applying the idea of authority to technologies may offer interesting avenues for analysing their compelling influence on what people think and do, it also poses new challenges and requires more conceptual clarity. How to understand this technological form of authority? Where does it come from and how is it constructed? While technologies now may be able to communicate, they still lack a mind, do not have intentionality as human beings do, let alone any form of freedom (Verbeek, 2011: 13). In other words, what would this mean for the authority of things?

1.7 RESEARCH QUESTION AND GOAL

The research question of this dissertation is:

> To what extent can digital technologies be perceived as authorities?

The goal of this book is to conceptualise a specific and compelling form of technological influence on human decisions, to further explore the significant extent to which human beings are willing to heed digital technologies, even beyond their own judgement, and to highlight its potential consequences. Based on the previous sections, I believe the idea of authority has a lot to offer for understanding interactions between human beings and technological artifacts and provides avenues for mitigating some pressing risks. It could help designers, for example, to better anticipate the authoritative roles of their designs and

the potentially harmful effects by modifying user guidelines, instructions, or interfaces. Or it could encourage policy makers to develop new regulations, should this be necessary.

For this, I propose to develop the idea of *digital authority*. For clarity's sake: digital authority, here, is not so much about the presence of human authority in the digital world. For instance, sometimes the term is used to indicate how online actors – such as social media influencers and websites – shape digital discussions and ideas (Vitali-Rosati, 2018; Casero-Ripollés, 2021) or to describe how traditional *offline* authority figures manifest themselves in cyberspace (Guzek, 2015). There have also been examples of how the internet has facilitated imposters with new ways to wield authority effectively. A case in point is the Pathé cinema chain falling to fraudsters in 2018. In a series of e-mails from the "head office", the management was requested to transfer money for a takeover in Dubai. Despite their doubts about the situation, the victims' trust in authority led them to transfer more than 19,2 million euros (NOS, 2018). While these are fascinating topics, this book moves beyond the authoritative communications of people *via* computers.

Neither do I intend to use the term to focus on digital technologies programmed to make autonomous decisions themselves in situations where people are absent. Think for example of a computer that governs the operations of machines in a factory (Kuflik, 1999), a fully autonomous weapon system that engages targets without human intervention or organisations that rely on automated decision-making for public service delivery (Van Eck, 2018). Despite the importance of this issue, in the context of this book, digital authority is not a synonym for the autonomous decision-making powers of a technology.[15]

Instead, with the term digital authority I refer to the authority of modern technologies themselves and their ability to generate output that compellingly influences human action and belief. Or to be more specific, it refers to the influence of digital technologies and their capacity to produce consequential communications - output that is accepted by people out of trust and respect, or their willingness to act *as if* this were so.[16] In the context of this book, the term digital authority applies to both a particular kind of influence that some digital technologies can exert over human beings, as well as to those technologies that are commonly considered as being capable of producing such influence. This approach allows for a more comprehensive investigation of a broad range of technologies and potentially applies to all utterance producing technologies – not just robots and advanced forms of Artificial Intelligence (AI). Furthermore, the term digital authority highlights that for understanding how technologies are perceived and responded to one must look beyond just algorithms. While it is technically true that algorithms are at their core, for understanding the construction of authority it does not help much to reduce digital technologies to the sum of algorithms as is done through the idea of algorithmic authority.

Digital authority does not necessarily have to be intentionally designed – as is the case with persuasive technology (Fogg, 2003). Surely, there are many examples of technologies implemented with the aim of gaining the unthinking compliance of their

users. However, there are also many instances in which this may not have been the goal, and people end up listening to technologies that were intended more as decision-making support rather than decision-making superior. As introduced in paragraph 1.5, authority is grounded in nothing so much as the voluntary recognition of those subject to it. Therefore, it is not so much about the question whether or not a technology was designed to be authoritative; it is about whether, for whatever reason, people recognise its authoritative position in their interactions and respond to it accordingly.

The idea of digital authority is also not limited to any realm or activity, and we may find digital technologies in authority roles everywhere. However, many of my references will be to the field of security where digital technologies are increasingly involved in substantive decisions that directly bear on the lives of human beings. Much of my academic work has focused on contemporary security challenges, and it is in this realm that the potential impact of digital technologies on the lives of human beings is particularly high. In the subsequent paragraphs I set the stage for the rest of this book. Paragraph 1.8 further describes the philosophical foundation of this book and its focus on security. In addition, I warn against the fallacy of interpreting this book as authority or techno-skeptical in paragraph 1.9. The final paragraph of this chapter (1.10) presents the main outline of the book and the chapters to come.

1.8 PHILOSOPHY FOR HUMAN SECURITY

This book is a philosophical exploration of digital authority. Philosophy provides theoretical insights and frameworks for reflection on the world around us. It is the science of looking at things differently: philosophers question assumptions, challenge paradigms and explore alternative perspectives. To understand human beings in rapidly digitising societies, this is highly necessary. We need new vocabularies to grasp many of the unfolding social experiences and behaviors. These tend not to develop in a vacuum, but in interaction with empirical life. Empirical observations and knowledge form an important foundation for my analysis. The empirical work of others allowed me to uncover patters and develop my argument. Philosophical or theoretical reflections, in turn, construct new objects or make visible new entities of study. This way it provides depth to empirical findings and my philosophical exploration here, hopefully, stimulates new empirical research projects aimed at further validating our knowledge. Philosophy and empirical sciences have a synergetic relationship.

A philosophy of digital authority also facilitates the design of constructive solutions for the risks associated with blind reliance on digital technologies. Especially in contexts where the risks of digitisation become existential for individuals, the need for a better understanding and appropriate responses is urgent. The field of security can be considered such a critical context that warrants our attention.

Security is a fundamental need.[17] For human beings to function and develop, being safe from danger and fear is an important prerequisite.[18] In classical liberal conceptions, the state is often justified through its ability to protect its citizens from hazards and threats that they would otherwise face. The alternative would be chaos or a return to a Hobbesian state of nature (Ashworth & Zedner, 2014: 7). In exchange for people's willingness to renounce their right to self-government and obey the law the state has the duty to protect their people and their property. Digital technologies offer new ways to fulfill this duty and tackle all sorts of security problems. In areas ranging from law enforcement to national security and fraud prevention, we now find large (coupled) databases, facial and automated number plate recognition systems, risk-profiling technologies, and expert systems being used to identify and respond to security risks as early as possible (Hildebrandt & Gutwirth, 2008; Broeders et al., 2017). These technologies hold many promises such as improved efficiency, objectivity, flexibility, and cooperation (Marx, 1995; Graham & Wood, 2003; Woods, 2009; Završnik, 2021) and it is therefore no surprise that digital technologies have become pervasive in the field of security.[19]

But the use of digital technologies also involves new challenges and risks. Think, for example, of biases and discrimination in algorithms (Eubanks, 2018; Noble, 2018), issues related to missing data (Peeters & Schuilenburg, 2018) and the inability of technologies to consider the broader context and examine all relevant aspects of a problem (Mosier & Skitka, 2009: 209). Digital technologies do not always provide good output, and its use can thus have serious consequences. In recent years, more and more worries were expressed, for example, about the use of digital technologies for weapon permit screenings (NOS, 2019), fraud detection (United Nations, 2019) and predictive policing (Amnesty International, 2019).

Much of the digitisation of security, however, remains largely invisible to the public. In some cases, this may be due to a general disinterest or incomprehensibility of the issue. However, as argued elsewhere, it is also due to "the secrecy that often shrouds security policy operations as well as the experimental nature of some applications and the understandable – though regrettable – reluctance to debate those in public" (Broeders et al, 2017: 311). This makes that it is also not easy to discuss how and to what extent digital technologies shape the field of security. Given the potential societal impact of this development this observation is worrying.

It is obvious, though, that digital technologies will continue to become more involved in human decision-making in and beyond the field of security, just like their potential impact on the lives of individuals and on societies will only grow further. This requires closer scrutiny as at least some of the problems associated with this development seem to arise from misperceptions and incomplete understandings of what both people and technologies do. The impact of technologies is often described by either referring to their intrinsic capabilities – what these technologies *can* do - or to the instrumental purposes of their users – what people *want* to do.[20] The risk, then, becomes that both technologies and

humans are approached in isolation, and are only understood in terms of their set potentials. But as philosophers have argued, technologies are *multistable* – they lack an inherent "essence" and instead derive meaning from specific contexts and modes of usage (Ihde, 1990: 144; Verbeek, 2005: 118). When a technological artifact, however well designed, passes through human hands, it gives rise to behaviors that its designers might not have anticipated and to behaviors that were not meant to occur. Such unanticipated interactions between human beings and technological artifacts can bring much welcomed serendipity but can also cause undeniably problematic scenarios (Coiera. 2007: 98; Tenner, 1997). To understand the role and impact of technologies in our world, it is thus important to focus on the technologies, their users, and the context in which they operate. If we do not, we run the risk of underestimating potential problems that come with technological innovation or overestimating the effectiveness of proposed solutions for dealing with these. And especially in the field of security this can have serious consequences.

1.9 NOTES ON AUTHORITY-DISCOMFORT AND TECHNO-SKEPTICISM

There are two issues to address before concluding this chapter. The first is about the negative connotation of authority and its associations with constraint, submission, and exploitation. Philosophers have extensively discussed the precarious relationship between *political* authority and freedom.[21] Horrific crimes and psychological research, such as that of Milgram, in turn, have highlighted the potentially grave consequences of obedience to authorities. And scholars such as Erich Fromm (1936) and Theodor Adorno (1950) identified the "authoritarian character" or personality as an explanation for the effective use of violence by ruling powers and as an answer to why people were attracted to fascism.[22]

These observations have contributed to authority often being associated with her excesses and to a general discomfort of anything that has to do with it. Such associations, some have argued, even resonate with some typical authority figures themselves who seek to avoid the term and manifestations of hierarchy in their interactions with others (Rood, 2013: 140-56). Yet while for some authority may be a contaminated concept, interpreted as potentially strict and unfree or mistakenly equated to formal power of some form (see paragraph 1.4), this does not do justice to all those instances where people willingly search for and need authorities to function and achieve results. The truth is that we would often feel lost without them. In this book, I do not intend to speak of (digital) authority as something negative or positive. Authority is a fact of life, and it comes with both chances and risks.

The second issue I want to raise here concerns my general position towards technology research and technological development. There is a variety of perspectives on technological development and its social consequences – as I also discuss in chapter 7. For instance, there is *techno-optimism* or *techno-romanticism*. For techno-enthusiasts scientific discovery and

technological innovation equals progress (Van den Berg, 2009). They embrace the idea that technological development will benefit individuals and society. Technologies are seen as the solution or a potential "fix" for many of our modern-day social and natural troubles. Conversely, there is the perspective of *techno-pessimism* or *techno-skepticism*, consisting of a more general critique or dystopian view of technology (Swierstra, 1997). The unprecedented growth of the technological society in the 19th and 20th centuries resulted in doomsaying among some, as it was accompanied by several rude surprises revealing technology's defects, including the looming nuclear war in the 1960s, the 1986 Chernobyl disaster and various environmental problems (Winston, 2014: 14). For techno-pessimists it became crystal clear that modern technology was not a blessing, but potentially dangerous in many ways. Instead of welcoming the technical fixes, they see in technologies a (potential) cause for many societal problems, as threat to our traditions and even to our species' future. As Bibi van den Berg writes, such interpretations often point out that "all of the promises of a better future that the techno-enthusiasts present – for instance, more autonomy, control, flexibility, and better means of communication – might also be understood quite differently: one could argue that the relentless and all-encompassing 'technological colonization' of our world contributes to a loss of control, an increase in inflexibility, and the diminishment of autonomy" (2009: 26). Winner (1978) describes that, for techno-skeptics technology has somehow "gotten out of control" and has become a source of domination that poses a direct threat to human freedom. In less extreme versions, techno-skepticism entails the belief that technological development brings with it problems such as social alienation, "choice stress and paralysis" and identity confusion which will, overall, degrade the quality of our lives (Verbeek, 2005: 35 c.f. van den Berg, 2009: 26).

Techno-optimism and techno-pessimism are often placed at the extremes of the same continuum. While they are prominent in the public debate, they provide us with superficial, one-sided analyses of technology as either good or bad. However, things are a lot more complex. As argued by Van den Berg (ibid.), technological changes typically result in a mix of both positive and negative outcomes. Different technological artifacts can have diverse effects in various contexts. Smartphones, for example, may change the postures of children, looking down at a tiny screen all day, causing serious back problems at a younger age. At the same time, a smartphone also offers new ways to stay healthy by providing us running schedules or recipes for low-calorie dinners. Therefore, we should be careful and try to refrain from taking either of these approaches as they potentially hinder our understanding of the role of technologies and technological development in our world.

With this book I aim for the pragmatic middle ground of the continuum (ibid.). I do not intend to depict digital authority as something to be feared or cherished. Instead, I strive to describe this development "as it is": ambivalent, as having both positive and negative effects that we must investigate in more detail. Such a *techno-pragmatist* view presupposes that we refrain from oversimplifications and transcendentalist descriptions of Technology.

As described by Peter-Paul Verbeek (2005:4), instead of studying "Technology-with-a-capital-T" - in its entirety, as a singular, separate, monolithic domain – we should focus on specific *technologies* and concrete technological practices. The idea here is that technology cannot be understood in isolation from social reality. Technologies actively contribute to human lives and practices (Verbeek, 2011: 5). Conversely, the form of technological artifacts and the role they play is the result of how they become socially embedded. Therefore, we cannot predict or study "from a distance" the effects of technological development, but rather need to take a closer look at technologies in a social context. This certainly also applies to the idea of digital authority that is central here.

1.10 OUTLINE OF THE BOOK

This book is divided into three parts. The first part is dedicated to setting the contextual and conceptual stage. Chapter 2 provides a rough sketch of our digitising world. This world is characterised by increasingly capable and connected technologies and an abundance of information. I discuss how these developments have impacted many human decision-making contexts. Among other things, it has resulted in technologies that could improve our cognitive functioning while at the same time cannot be blindly relied upon.

In chapter 3, a basic ontology of authority is provided. As already touched upon above, authority has many meanings and refers to a diversity of actors - from allegedly wise experts to those holding formal offices such as police officers and judges. This chapter builds on the preceding discussion and what they all seem to have in common; a capacity to issue communications that motivate its recipients to respond with attitudes of trust, respect, docility, and acceptance. But how is such compelling influence constructed? For the analysis of authority, I identify three conceptual layers: the relation, the situation, and its cultural context. They form the starting point of this book's investigation into digital authority.

The second part of this book seeks to further explore the idea of digital authority. Chapter 4 focuses on digital authority relationships. What would they look like? How are they structured? The work of Don Ihde (1990) provides important initial insights here on the different ways in which human beings can relate to technological artifacts. I suggest considering digital authority as a specific human-technology relation. This relationship hinges on technologies' ability to communicate as well as on the experienced asymmetry that characterises all forms of authority. Subsequently, this chapter revolves around some of the technological qualities that could support this relational asymmetry and the capacity to be consequential.

Chapter 5 is about digital authority situations. Using Erving Goffman's notion of "definition of the situation" and the concept of scripts I suggest that digital technologies can incite people to understand their interactions with a technology as an authority situation

in which their attention, compliance and acceptance is warranted. To describe why this is the case I turn to the literature on *anthropomorphism* and the *Computers are Social Actors* (CASA) paradigm. Both lead us to conclude that when digital technologies begin to resemble human authority figures, it becomes quite likely that people respond to them accordingly.

In chapter 6, I turn to authority's cultural layer. It focuses on the conditioning of attitudes and expectations towards authority and digital technologies as part of contemporary cultures. I describe two broad, cultural perspectives on technology: one in which technologies are much celebrated, the other in which technologies are a serious source of concern. Furthermore, following Anthony Giddens (1990), modern culture is characterised by a constant demand for knowledge and a search for reliable others. Could this make a fertile ground for digital authority?

Up to this point this book's focus has been primarily on the construction of authority. The third part, in turn, discusses the limits of digital authority and situations wherein people say no to a technology. In chapter 7 I will be looking at authority relations, situations, and cultures again to identify various grounds for the rejection of digital authority. The chapter ends with a reflection on the question how people can be equipped to deal with digital authority and its potential risks.

Chapter 8 concludes this book, condensing its main points. The final sections consist of an epilogue, discussing some issues that deserve further thought and future research.

1