Computer says no: understanding digital authority

Weggemans, D.J.

COMPUTER SAYS NO

Understanding Digital Authority

Daan Weggemans

This dissertation was supported by the municipality of The Hague.

Cover image: Frans Martens
Layout and print: Ridderprint | www.ridderprint.nl

Computer Says No
Understanding Digital Authority

PROEFSCHRIFT

Ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op 13 februari 2025
klokke 16:00 uur

door

Daniël Jochem Weggemans
Geboren te Amersfoort
in 1986

Supervisors

Prof. dr. B. van den Berg
Mr. dr. Q.A.M. Eijkman

Doctorate Committee

Prof. dr. D.W.J. Broeders
Prof. dr. C.L.J. Caminada
Prof. dr. V.A.J. Frissen
Prof. dr. E.L.O. Keymolen (Tilburg University)
Prof. dr. S.L. Kuipers

An old man enters a hospital and walks up to the reception desk to register.

"I have an appointment to see the cardiologist. My name is Peter Bale."

The receptionist types on a computer, looks up and replies:

"Computer says no. Apparently you are already dead."

"That must be a mistake" the man says.

The receptionist looks at the computer again.

"Dead" she nods affirmatively.

Hopeless and bewildered the man left.[1]

1 Little Britain Sketch 'Computer says no: 14'.

CONTENTS

AUTHOR SAYS

We are shaped by the interactions with others and the settings we find ourselves in. This also holds true for the things we make. In any artifact - whether tables, phones, or dissertations – lie hidden the ideas, criticisms, passive aggressive sighs and tons of time, patience, and support of others. Nothing happens in isolation. This book developed out of the countless encounters with people that have inspired and supported me over the years.

I owe a lot to my PhD supervisors, prof. dr. Bibi van den Berg and mr. dr. Quirine Eijkman. I have been lucky to work with true authorities like you and hope we will continue this in the future. My discussions with Bibi have been such a source of inspiration. She has a flair for exciting conceptual ideas, asking the right questions and sharing timely observations. It has truly enriched this dissertation in many respects. Bibi has been a true mentor and pivotal in my career. I am deeply grateful for her encouragement and belief in my abilities. Quirine provided me with the opportunity to work on this project and remained important throughout. Even across different organisational settings, she has always been there to support me whenever needed. I have greatly benefited from all the constructive feedback that Quirine has given me. Her advice on many occasions has allowed me to develop as a scholar and find my own place in academia. I also want to express my gratitude to prof. dr. Edwin Bakker and dr. Ruth Prins, whose reflections and motivations have been instrumental since the beginning of this project. They generously took the time to discuss early drafts and provided helpful comments to enhance my arguments.

I am also grateful for the resources made available by the city of The Hague. It has allowed for a study that goes beyond the desires for quick returns and overly ambitious solutions. Doing fundamental and conceptual research is an important way of acquiring new knowledge and the basis for innovation in the long run.

While my academic foundations were established in Nijmegen, I am excited to reflect on the start of my academic career at the Centre for Terrorism and Counterterrorism (CTC) and the International Centre for Counter-Terrorism (ICCT) in The Hague. I had the opportunity to work with and learn from Beatrice, Jaap, Bart, Jelle, Liesbeth, Eva, Orla, Wietse, Peter, Eelco, Sergei, Tinka, and many others. I hold dear many of our conversations that have certainly found their way into this book.

Later, as part of the Institute of Security and Global Affairs (ISGA) I entered a stimulating environment with people from many different disciplinary backgrounds. I thank Sanneke, Jeroen, Koen, Marieke, Dennis, Willemijn, both Pauline's, Ernst, Layla, Tommy, Els, Celesta and all other colleagues at the institute and faculty for their support and ideas. Specifically, I also thank those in the Cybersecurity Governance Group for their input and feedback on this book. As programme director of the Bachelor Security Studies, I felt much appreciated in the educational team at ISGA and the faculty of Governance and Global Affairs. It was a privilege to work with so many talented lecturers, tutors, and coordinators. You made writing this book (a bit) easier.

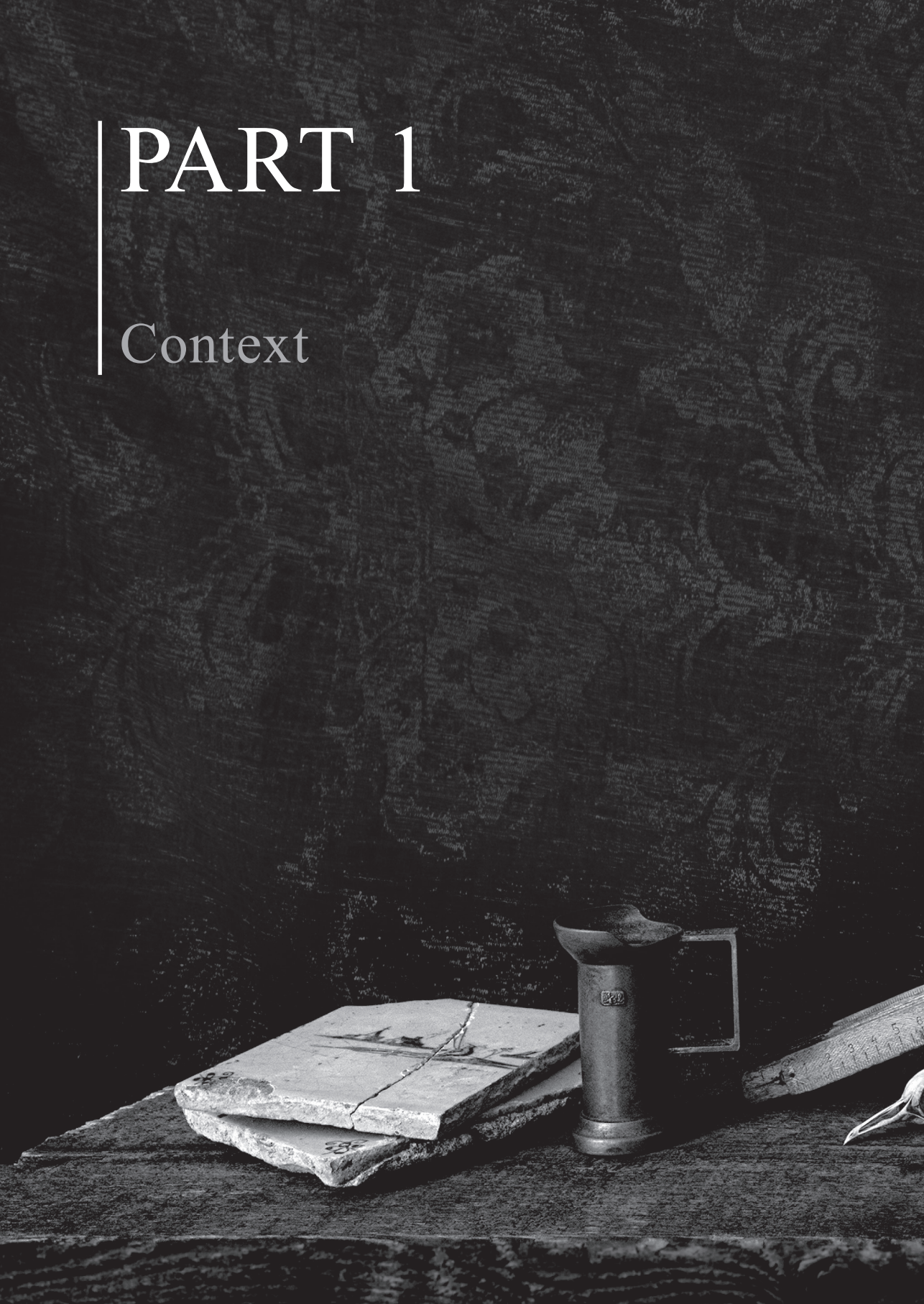
It is my strong conviction that the integration of science and practice is of great importance. "Going out there" provides us with a perspective on what really matters and allows us to make a difference. Over the years I had the opportunity to visit many organisations and meet many experts and professionals in the field of security. Through them I learned about the role and complexities of digital technologies in security practices. It allowed me to place my philosophical reflections in empirical realities. I am also grateful to the *Landelijk Steunpunt Extremisme* (LSE), where I collaborate with so many committed professionals and learn invaluable lessons about what human well-being and security truly entail.

There are also people who I have never worked with but have been central in writing this dissertation anyway. I am surrounded by sweet people who divert my attention away from the overwhelming with impromptu drinks, movies, and dinners. Thank you to all of you who have cheered me on and nodded along whenever I explained my research.

The final words are for my family. My parents and sister gave me roots and wings. They encouraged me to explore life and supported me unconditionally while doing so. I am deeply thankful for this. I am also lucky to have not just a great uncle, Frans, but an uncle who was willing to make the wonderful photo on the cover of this book. I would like to dedicate this dissertation to Ria, Ise and Maxim. With you I experience every day that love is the best kind of influence. No authority can match that.

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

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.

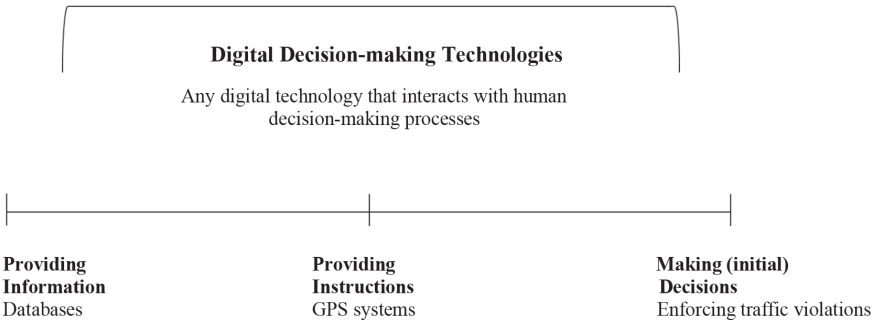


Figure. 1 Digital technologies and human-decision making

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.

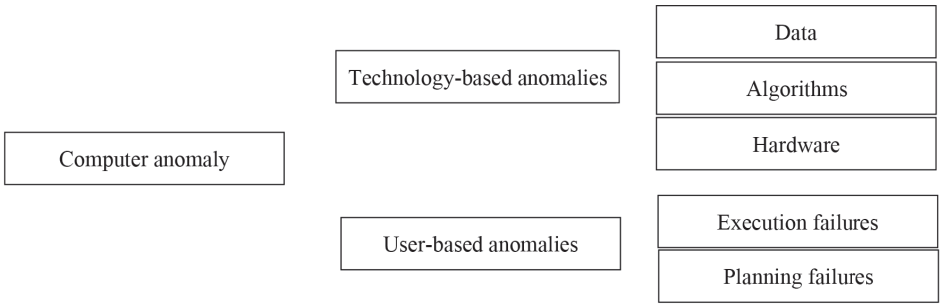


Figure 2. Understanding information system anomalies

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.

2

3 AUTHORITY

What do we mean when we speak of authority? How does it function and where does it come from? The aim of this chapter is to reach a better understanding of the concept of authority. Despite it being used prolifically, there is still much confusion over what authority means. The concept of authority is pervasive and yet elusive (Bell, 1970: 190). One reason for this, as Carter (1979:1) argues, is the ambiguous use of the term. Take, for instance, the sharply diverging connotations of two adjectives associated with authority: authoritative (to indicate that something is likely to be accurate or true) and authoritarian (referring to a concentration of power and dictatorial rule). Authority is also used in very different contexts, ranging from the realm of politics to that of knowledge and from the family and education to the military and organisational life, making it difficult to pin down its exact meaning.

Generations of scholars from various backgrounds have worked to clarify the nature and workings of authority. Social psychologists have been concerned with the strong human tendency to comply with authority figures (Milgram, 1974). Political philosophers have had fierce debates regarding the perceived crisis of authority that society has been facing ever since the beginning of modernity (Arendt, 1961), and the possibilities for reconciling authority with autonomy (Raz, 1990; Wolff, 1990). And sociologists and public administration scholars owe much to Max Weber's (1919) engaging discussion of the development of modern bureaucratic organisations and the subsequent shift from the ideal types of traditional and charismatic types of authority to that of rational-legal authority.

In both scholarly and public discussions authority tends to be "thingified". That is, authority is often depicted as a substance that is simply possessed by certain people who use it to influence others. I consider this ontological presumption to be unsatisfying as it excludes the interactional dynamics through which authority is constructed - and deconstructed. As sociologist Richard Sennett concluded: "authority is not a thing. It is an interpretive process which seeks for itself the solidity of a thing" (1981:19). I argue here that authority cannot be reduced to detached people, institutions, or *digital things*. This chapter presents a layered understanding of authority as an asymmetrical relationship (paragraph 3.3) that is essentially situational (paragraph 3.4) and contextualised by culture (paragraph

3.5). Based on these insights, at the end of this chapter, a general idea of digital authority is introduced. The last paragraph summarizes the main conclusions of part 1. But let us first begin with the word authority itself, at its etymological roots (paragraph 3.1).

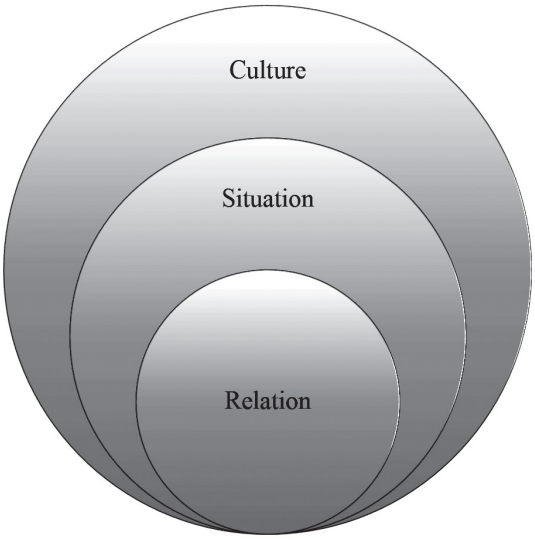


Figure 3: The construction of authority

3.1 AUCTORITAS AND AUTHORITY

Today, the meaning of authority has become blurred. For one thing, this is because the term is often used interchangeably with power. For example, saying that someone has "the power to do something" often means the same thing as saying that this person has the "authority to do it". But as is discussed later, there are important differences between power in its coercive sense and authority. An initial step to re-establishing some of these nuances is returning to authority's etymological roots.

The word authority is derived from the Latin auctoritas, a significant term in various spheres in ancient Rome ranging from family life to politics to economical transactions. While this demonstrates its importance throughout Roman society, the frequency and various functions of the term also make it notoriously difficult to define, as Grant has put it (1969: 443). The noun *auctor* is at the root of *auctoritas*. Like the English *author*, an *auctor* is an originator - someone who creates, inspires, invents, promotes, or produces something (Watt, 1982: 11; Friedman, 1990: 74-6). Examples of people who were called an *auctor* were parents, ancestors, and givers of advice. It is said that their role of creator indicates special access to, or some form of power over an object and were thus regarded as superior actors

(Pruitt, 2011: 23). *Auctoritas* and *auctor* are also linked with the lost verb *augere*, meaning to initiate, to augment, to enrich or to increase (Arendt, 1961: 121-2; Furedi, 2013: 61). In a family, for instance, the oldest known forefather - the founder - was called its auctor, and his *auctoritas* was inherited by his descendants. Each generation augmented and increased what the initial founder had begun (Watt, 1982: 11-2). In a similar vein, Rome itself was also considered having an auctor. Watt argues that it was the foundational act of Romelus that "was the source of the civil *auctoritas* of those who, in unbroken succession, were to share his work as an auctor in time to come, by augmenting the Roman inheritance which in their turn they passed on to their successors" (ibid).

In political life of republican Rome, *auctoritas* did not encompass a formal right to rule. On the contrary, it was directly contrasted with *potestas* (related to the English word power) and *imperium* (military power) - the formal rights of magistrates to issue lawful and legally enforceable commands (Watt, 1982: 12). The Roman Senate, however, did not have such a formal right to demand obedience, they only had *auctoritas*. It could not issue commands backed by the force of law. In fact, the Senate did not command at all, but only provided civil officials with executive powers with (authoritative) advice. As April Carter observed, the Senate could, for example, call for a respectful hearing and because of the collective *auctoritas* of its members, "the popular assembly would usually do what the Senate urged" (1979: 17). Thus, while the Senate had no executive power on its own, its deliberations and advice often did make public policy. Krieger elaborates that while the Senate was "just an advisory body [of] fathers or elders" with neither legal status or compulsory force but only *auctoritas*, it was, in fact, the Senate that "bound the Roman citizens to compliance far more definitively and extensively than did the other offices of the republic" (1977: 25). In an attempt to capture the Senate's *auctoritas* Carl Friedrich described it as "advice which cannot be safely disregarded". (1972: 47) and Mommsen (1887 c.f. Arendt 1961: 123) as being "more than advice and less than command; an advice one may not safely ignore".

Unlike the formal powers vested in the office or position of a magistrate, the scope of *auctoritas* in the political sphere was not clearly defined. According to Wirszubsi (1950) the fact that *auctoritas* was not constrained by formal limits had far-reaching practical implications. He writes that, "whereas *potestas* is confined within certain limits, there is, at least in theory, no limit to the scope of *auctoritas*: it can be brought to bear on any matter" (c.f. Furedi, 2013: 63). The promise of a claim to such far reaching influence, Frank Furedi concludes, made that many leading Romans vigorously promoted their *auctoritas* (ibid). The best example is the first Emperor of Rome, Augustus, who tried to legitimise his hegemonic rule by referring to his supreme *auctoritas*. In *Res Gestae Divi Augusti* ("The Deeds of the Devine Augustus"), a document Augustus wrote at the end of his reign, he says:

> "In my sixth and seventh consulships [28-27 BC], after I had extinguished civil wars, and at a time when with universal consent I was in complete control of affairs, I transferred the republic from my power to the dominion of the senate and people of Rome. For this service of mine I was named Augustus by decree of the Senate, and the doorposts of my house were publicly wreathed with bay leaves and a civic crown was fixed over my door and a golden shield was set in the Curia Julia, which, as attested by the inscription thereon, was given me by the Senate and people of Rome on account of my courage, clemency, justice, and piety. After this time, I excelled all in influence [*auctoritas*], although I possessed no more official power [*potestas*] than others who were my colleagues in the several magistracies."[52]

Augustus, self-consciously, contrasted his claim to influence with the formal office and legal powers that both he and his colleagues in the magistracies possessed. *Auctoritas*, in this sense, represented a claim to something more than the office of leadership, Furedi concludes (2013: 63). Augustus understood that such a claim was difficult to dispute and that *auctoritas* was "the kind of substance on which real influence [was] based" (Galinksy, 1996: 15).

Again, the Roman use of *auctoritas* encompassed a much wider scope than that of politics. Among other things, the term also referred to the more general communicative quality of individuals who "spoke with authority" or were thought of as "moral authorities" (Furedi, 2013:6). The old and the wise and those with an extensive record of achievement were attributed *auctoritas*. Cicero, for example, spoke of the "*auctoritas* of old age" (Galinksy, 1996: 16). Those who possessed such a quality were entitled to speak and to be heard. Similarly, knowledge (*scientia*) itself was also said to have *auctoritas*. And a*uctoritas* was attributed to priests, bishops and to the Roman gods themselves.

Furedi writes that *auctoritas* initially functioned as a (pre-political) term which expressed "the capacity to hold responsibility for a variety of relationships and transactions" (Furedi, 2013: 61). *Auctoritas tutoris*, for example, refers to the pronouncements of a trustee or guardian, *Auctoritas patris* to the approval by the head of the family (*pater familias*) and *auctoritas venditoris* was called upon as guarantee for the closing of contracts and important sales (Krieger, 1977: 258-9). In this context, *auctoritas* encompassed communications that would move things forward.

Auctoritas has been labelled as the key to capture the distinct nature of authority (Peters, 1958: 210). Yet, as illustrated here, its many different shades of meaning complicate the construction of a strict definition. What can be concluded, however, is that *auctoritas*, in its many forms, always includes a claim to respect, trust and esteem that exerts a compelling influence on others, but remains distinct from force (Furedi, 2013: 62). With *auctoritas*, whatever an actor says or does, is taken to be right and to be listened to.

3.2 CONSEQUENTIAL COMMUNICATIONS

Contemporary usage of the term authority is perhaps as fuzzy and ambiguous as its Latin original. A useful distinction that is frequently made is between being "an authority" in the realm of knowledge (also called *epistemic* or *cognitive authority*) and being "in authority", pointing at the office or position which formally entitles someone to make decisions about the behavior of others (Peters, 1958; Flathman, 1980; Wilson, 1983; Friedman, 1990; Pierson, 1994; Kruglanski et al., 2005). Richard De George (1985) adopts a similar division. On the one hand, he speaks of *executive authority*, referring to the exercise of authority by someone (or an institution) in the capacity of holding an official "right or power to command or act for or on someone else" – such as police officers, political leaders, and judges (De George, 1985: 22). On the other hand, he speaks of *nonexecutive authority* which does not include such an official right but does allow for the influencing of others. An example of *nonexecutive authority* is an *epistemic authority* who is thought of as having special knowledge, wisdom, or insight (e.g., an expert, a scholar, a technical specialist, a social media influencer, etc.). Furthermore, this type of authority also includes those who serve as a practical example for others (such as the artist who sets the standard for artistic practice) (ibid: 23).[53]

To transcend these distinctions, I have found the work of Bruce Lincoln (1994) particularly helpful. He highlights that with authority, someone can bring about a discursive effect. Authorities have the capacity to produce *consequential speech*; the kind of speech that quells doubts and wins over the trust of people with whom they engage, Lincoln writes (1994: 4). The French philosopher Bertrand de Jouvenel (1957) expressed a similar view, speaking of authority as "the ability of a man to get his own proposals accepted" (1957: 30-1) and theologian Austin refers to it as the "capacity to get things done" (2010: 21). While there may be differences in how people come to occupy their positions and the warrants they can produce in support of their authority, Lincoln argues, what all authorities share is that they can perform consequential speech acts - therewith also connecting authority to auctoritas' decisive influence discussed above. As an example, he points at the words of executive authority in its most extreme form - that of military command - which generates automatic and unquestioning obedience, in the same way as the utterances of an epistemic authority will move a debate beyond dispute (Lincoln, 1994: 4).

It makes sense to expand this view beyond the medium of language. Authorities do not always need words to yield results. In fact, a disapproving glance, a sigh, or a benevolent nod can work just as well. Hence, I would rather speak of authorities' capacity to produce consequential *communications* – communications that motivate an audience to respond in a way that signifies acceptance. Following Lincoln, I take the consequentiality of such authoritative communications not so much to depend on what is communicated or on how it is communicated. Authorities do not necessarily need to express themselves particularly well to gain compliance. Neither does this automatically flow from individual qualities such as office, wealth, ancestry, or charisma. Instead, Lincoln says, authority is a relational outcome,

grounded in a perceived or institutionally ascribed asymmetry between an actor and his audience that secures attention and produces attitudes such as trust, confidence, respect and even reverence – or the willingness of people to pretend this were the case (Lincoln, 1994:4). When people recognise this relational asymmetry - whether through notions of legitimate institutional representation or through demonstrated exceptional characteristics inspiring trust and respect - they are inclined to adhere what is communicated to them. In the following paragraphs I further substantiate this by making explicit three fundamental premises that form a framework for the further analysis of authority.

3.3 AUTHORITY IS RELATIONAL

Authority is not an intrinsic quality. It is not first "possessed" by someone and then "used" – even though we tend to conveniently put it this way. We often speak of someone – or as explored in this book, of something – *having* authority, but we cannot capture authority in this material sense. Authority is relational. It only comes to exist when we enter a relation. When there is no one to listen, there is no authority. Just like there is no authority when there is no one who speaks or communicates in some other way. There is only authority when a speaker is listened to, and an audience is spoken to (Bingham, 2008: 9).

As mentioned above, authority implies an asymmetrical relationship, in which one is recognised by the other(s) as, in some sense, a superior, not an equal. It is this asymmetry that is constitutive of the capacity to produce communications that are adhered to. Because of authority's discursive and asymmetrical character, it is often associated with two other forms of influence: persuasion, and coercion – the processes in which someone is convinced or forced to do or believe something (Lincoln, 1994: 5). But, as many scholars who have tried to capture the nature of authority have argued, authority remains distinct from persuasion and coercion alike. Peters, for example, writes that authority applies to those situations in which "conformity is brought about without recourse to force, bribes, incentives or propaganda and without a lot of argument and discussion" (1958: 218). Authorities can compel obedience that is simultaneously free and unquestioned (Arendt, 1970: 45).

With authority, there is no need to persuade or coerce to make someone think or act in a certain way. Take the simple example of a mother telling her son to do a chore. With authority, the son will do what he is asked to do simply because it was his mother who told him to do it. She does not need to persuade him – using all sorts of arguments to convince that doing the chore has all sorts of advantages. Such rhetorical strategies presuppose a certain equality between both. Similarly, she also does not need to threaten or use force to get the child to do it.[54] The son simply follows her words. Indeed, the mother may very well be able persuade or force her son. But when there is authority, the chore will be done without her having resort to argument or force. Surely, as I have learned from personal experience, there are many instances in which children do not listen to their parents, and children's

rooms often remain messy despite repeated requests. This can indicate the (temporary) absence of authority. Parents and other authorities can become quite preoccupied with finding appropriate responses to such challenges to remain consequential. Nevertheless, the general idea is that under the sway of authority, people will end up doing as they are told.

"Authority has the notion of acceptance built into it", Watt writes (Watt, 1982: 30). Authority relations are recognised voluntarily by their participants. Richard Sennett captures this dynamic and speaks of authority as a "*bond* between people who are unequal" (Sennett, 1981: 10, emphasis added). And because this is the case, authorities can gain compliance without having to threaten or argue. On the contrary, Lincoln writes, the exercise of authority may rest on the "naked assertion that the identity of the speaker warrants acceptance of the speech" (Lincoln, 1994: 5). In his best-selling book, *Influence*, Robert Cialdini fittingly refers to authority as "directed deference" (Cialdini, 2007:208). When people recognise authority, it is said they are convinced in advance that whatever is communicated to them is worthy of respect and acceptance. As Friedman (1990: 64) has described it, authority relations involve some degree of surrender of judgement – when people recognise an authority, they will accept his or her communications regardless of their own examinations and evaluations.[55] The fact that an authority tells someone to do or believe something, is enough to motivate this person to follow – sometimes in an automatic or blind fashion.[56] What matters for them is not the content of what is said but rather the source - who says it. This is frequently exemplified by someone saying that he or she did something "simply because X told us" (Peters, 1958: 218).

While we must distinguish authority from the modalities of coercion and persuasion and reserve the word authority for situations of unforced and un-persuaded compliance (Watt, 19982: 29), both are also constructive elements of the asymmetry that is endemic to authority relationships (Lincoln, 1994: 5-6). Carl Friedrich, for example, has argued that authority hinges on the "capacity for reasoned elaboration" (1972: 52-5). According to this idea, it depends on the existence of good reasons for compliance and that these reasons can and will be explained by authorities when asked to do so (ibid: 55). This can also be linked to the earlier discussion of an *auctor* having "differentiated access" to something. Authorities are often perceived to have direct access to certain experiences or knowledge from which their subjects are barred (Friedman, 1990: 75).[57] In practice, it is more about individuals *believing* or *presuming* that such capacity exists. When a doctor says, "take these pills three times a day", a common response is to follow this advice. We do so because he appears medically well-trained, and we assume he can give us elaborate reasons for his advice if asked (and circumstances permit). The doctor may be very well an imposter or a charlatan: if his patients trust that convincing reasons can be given for his utterances his authority is real in its effect. In this view, persuasion then lurks in the background of authority as a presumed possibility that can be brought forward at any time. In a similar vein, authority is related to the capacity to use force. As part of a larger hierarchal social

system, some actors are delegated special access to coercive power. This can play a role in the asymmetries between teacher and student, parent and child or soldier and officer where everyone is aware that under certain conditions - for example in the case of observed resistance - authorities have a right to use force rather than reason. Hence, the belief that under challenging conditions force will be deployed can also underpin an individuals' regard for a speaker and their willingness to accept authoritative communications.

In short, persuasion and coercion are supportive parts of authority relations, Lincoln emphasises (Lincoln, 1994: 5-6). By being present in the background, as possibilities that lead an audience to assume that there must be good reasons or dire consequences for their deferential responses, they co-shape authority's relational asymmetry. But only in a state of latency this holds true; their actualisation indicates authority's very absence (Lincoln, 1994: 6). Once authority must be self-consciously justified this relationship is suspended - at least for the time being. At those moments, when authorities might sense they are losing the trust and respect of their audience, they can become tempted to rely on coercion and persuasion. But, as Lincoln writes, when authority operates on pain and fear or through arguments, it ceases to be authority and becomes (an attempt at) persuasion or coercion, adds Lincoln (Lincoln, 1994:6). Ultimately, this can lead to the definitive deconstruction of authority. As Arendt has put it:

> "A father can lose his authority either by beating his child or by starting to argue with him, that is, either by behaving to him like a tyrant or by treating him as an equal. To remain in authority requires respect for the person or the office. (1970: 45)."

3.4 AUTHORITY IS SITUATIONAL

Authority is not just relational but also grounded in the character of the situation: it is the impression conveyed to an audience in a particular setting that incites its acceptance, deference, or obedience. As such, authority always involves a theatrical element and the use of discursive practices that impose a reading of the situation on others. Authorities rely on an array of theatrical instruments to foster the right impression among their audiences. This view also aligns with Erving Goffman's dramaturgical approach. Goffman writes that everyday interaction is a theatrical performance in which individuals convey and maintain impressions of themselves.[58] To influence others, people put up a "front" and present an idealised image, strategically using certain language, gestures, symbols, and the setting in which they perform (Goffman, 1959: 32-40). The things that might undermine such "frontstage" acts are carefully kept "backstage", out of view of the audience.[59] With their performance, Goffman argues, individuals aim to express and maintain a definition of the situation – they make suggestions to others how to define what is going on and how

to respond accordingly (Goffman, 1959: 32). Authority, from this perspective, is thus not so much about solid content but about the creation and maintenance of perception. To command an audiences' trust and respect, one must make an impression that shapes (or defines) the situation as an authority situation, calling attention to the relational asymmetry that exists between the actor and the audience. This way it also ties in with Sennet's observation that authority is based on *images* of "strength and weakness" (Sennett, 1981: 4). Or as Niccolò Machiavelli famously said the prince "need not actually have all the qualities I have enumerated, but it is absolutely necessary that he seems to have them" (Machiavelli, 1513: 57).

To stage an authoritative performance, people may rely on a variety of dramaturgical practices and props. As De George writes, authority has many symbols - clues an audience may find in someone's appearance and performance which mark the authority of its bearers and function as a mental shortcut to identify authority. Sometimes these symbols are subtle, and sometimes more obvious (De George, 1985: 100-4). The classical example of such tangible symbols of authority are the uniforms and insignia of rank in the police and the armed services. Their apparel communicates authority, both within the organisation as well as to the public. Without their badges or uniforms, we would not be able to recognise them and are less likely to follow their pronouncements. This latter point is echoed in the seminal work of Leonard Bickman (1974). In a series of experiments Bickman showed how his research assistant on the street managed to ensure compliance of random pedestrians by wearing a police-style uniform. When the research assistant dressed differently - for example wearing casual clothes or a milkman uniform - compliance was much lower. Other studies also showed that the condition of a uniform and equipment can impact authority. For example, when a uniform is in a bad state people can become reluctant to obey or - as interviews with prison inmates who murdered police officers indicated (Pinizzotto & Davis, 1992) – these uniforms can even be an open invitation for resistance as they may communicate incompetence. What these and other studies illustrate is the powerful situational impact of a uniform and the way they convey the authority of the person wearing it.

Throughout history many more items have been used to designate certain people as distinguished for their office, status, or rank. Consider, for example, the crown and jewels of a king, the diplomas and awards of a scholar, the choir dress and special vehicle of the pope, the suit and tie of a corporate executive, the robes and wigs of lawyers and judges, the lab coat and stethoscope of a doctor or the chair and megaphone of a movie director. These and many other items all physically set someone off as special (De George, 1985: 102). They draw attention to the authority of their bearer in a certain situation and incite acceptance. Or as Lincoln concludes about this variety of uniforms and insignia: "insofar as a particular garment or piece of paraphernalia has come to enjoy the respect and reverence of a given audience (for whatever reasons), it focuses the attention of that audience on whomever comes to hold it, whose words will consequently be received - initially at least

- with similar respect and reverence (Lincoln, 1994: 8). Even when the bearers of these symbols of authority are not genuine these uniforms and insignia obtain for their bearers the trust of others and yield them de facto authority - such as illustrated by Bickman's research assistant pretending to be a cop.

Along with the tangible symbols of authority come the intangible ones. An example is the range of physical postures, characteristics, gestures, and expressions that function in concert with the visible items of authority.[60] Done at the right moment, a hand gesture, a glare, a frown, or a bow becomes a vehicle for conveying attitudes of decisiveness, trustworthiness, wisdom and thereby help to lay claim on people's attention and acceptance. To this it must be added that the relatively fixed aspects of physical appearance (e.g., racial characteristics, age, gender, size and looks) have, throughout time, also been the very reason for attitudes of distrust, envy, suspicion, doubt, wariness, and skepticism. This illustrates that authority is, in fact, dynamic, and cannot be understood insolation from its subjects and the setting in which it operates.

Another set of intangible symbols of authority is more verbal in nature. Language plays an important role in the construction of authority. As De George argues, "the language of epistemic authorities, even when obfuscating, sometimes gains them increased respect; and the language of executive authorities, when peremptory, decisive, and insistent, frequently wins them recognition, compliance, respect and obedience" (1985: 103). The right use of language is both a symbol as well as a basis for getting and keeping authority. Every field of knowledge has a specific vocabulary, and when this vocabulary is used in a natural way it distinguishes a person as a part of an inner circle and confers authority on him or her (ibid). Consider, for example, the vocabulary of many corporate executives, doctors, academics, lawyers, and sports coaches. Their language frequently brings with it a certain inaccessibility. Such incomprehensibility may be perceived as a sign of expertise or knowledge from which one is barred and can be a good reason for acceptance of someone's words.

De George also discusses the role of the various forms of address and titles. The use of titles marks certain people as authorities and generates distance between those who have a title and those who do not (1985: 103). Titles can be formal such as "Professor", "General", "Priest" or "King" but can also be more informal as in "Expert" or "Specialist". The psychological importance of these titles is that when they are used, they remind us of the asymmetry - whether in knowledge or in official position - between those involved.

Certain designated times and places can also be considered symbols of authority that predispose many people to accept the words of a speaker. Lincoln gives the examples of the churches' pulpit on Sunday mornings, the lectern at the appointed hour for a (university) class, and the judge's desk when court is in session (1994: 8-9). As he explains, these times and places are authorised not only by specific institutions in which people trust (church, university, government), but also "by the memories, associations, assumptions, and

expectations they call forth in the audiences they call together" (ibid). As such, times and places confer authority on those who are allowed to perform within these privileged settings.

Last, there is the role of others in shaping authority situations. For instance, individuals such as an office director, a school principal, a head coach can draw attention to a specific actor, signaling to an audience to listen actively and show respect to whoever they introduce to the stage – whether it is a newly hired manager, a substitute teacher, or a team captain. The presence of such henchmen can be an important step for constructing authority. Or take the actions and testimonies of others in the audience. When we see or hear others relying on authority, or rejecting it, we are more likely to do this as well – the idea of *herd instinct* (De George, 1985: 97). In line with this is the idea that perceptions of competence are often co-constructed by the testimonies of others in the audience. We frequently take reputation as a ground for our preliminary attentive responses.[61]

The significance of signs, symbols, and the role of others for authority is often underestimated. They vary across time and place. However, their general function remains the same. They create the impression that those associated with them are worthy of trust and respect and, as a result, influence the capacity to evoke compliance. As such they are important as they invite recognition of the authoritative position of the performer from the audience (and thus the asymmetry of the relationship). Or more in Goffman's terms, the right expressive equipment, places, and people may leave an impression on an audience that invites them to define a situation as an authority situation in which their compliance is warranted.

3.5 AUTHORITY IS CULTURAL

The view that emerges from these sections is that authority, above all, depends on the recognition and acceptance of those subject to it. Yet the audience itself is seldom a central topic of analysis. People observe, interpret, and respond in different ways to authority claims and this is strongly shaped by the set of culturally learnt values and beliefs nested in their minds. How we make sense of our world and the way we respond to the situations we encounter is co-shaped by our cultural backgrounds. When growing up, people in the same culture learn about basic assumptions, attitudes, symbols, heroes, and rituals, which together form a culturally specific meaning system (Hofstede et al., 2010; Meyer, 2024). In our interactions with others, individuals – often unconsciously – draw from this cultural repertoire of thinking to determine what is going on and what actions are deemed appropriate here.[62] In this sense, culture is also important for shaping our definitions and ideas of authority.

Learning to recognise and respond to authorities is part of every culture. From a young age, children are taught who to trust and respect. Within the family, and at school, children learn about the benefits of listening to proclaimed authorities and of the costs of doing something else. Later, people enter public life and work organisations, acquiring

necessary information about all sorts of professional roles and practices. They are told who to listen to and whose words to accept. This way, individuals' orientations, and expectations of authority are inextricably intertwined and emanate from within cultures and social communities. The symbols, rituals, and places discussed in the previous paragraph which allow speakers to convey authority, are often visible to everyone. However, their meaning and performative powers lie in the interpretations by cultural insiders. Over the course of their lives people develop specific patterns of thinking that enable them to recognise rules and behavioral standards, the positions that are designated as authoritative and the meaning of authority symbols, rituals, and places within their cultural context. Because this learning happens constantly, and largely unconsciously, following authorities may simply become a habit.

Authorities are, thus, often a reflection of a series of culturally conditioned expectations of an audience. It is the audience that, within a given cultural context, authorises a given actor to communicate with authority. It does so based on the things it has learnt from its environment. Hence, authority can thus differ per time and place. Authority is clearly not a given, as argued above, nor is it stable and discrete (Carlson, 2017: 12). Referring back to the preceding examples, the influence of the Roman Senate dissolved by autocratic emperors, anti-vaccination campaigners and naturopaths deny the authority of medical doctors, scientists are questioned by politicians and religious fundamentalists, the military and police is openly challenged by rioters, and so on. Authority is always open to contestation and change. With changing cultural horizons, recognised authorities may be forced to do new things to maintain the trust and respect of their audience, just like new authorities may emerge. While we no longer grow up believing in magicians and sorcerers, we may now discover cultural elements in support of new authority claims – by digital technologies perhaps?

3.6 TOWARDS A THEORY OF DIGITAL AUTHORITY

There are many volumes available on authority and I do not pretend this chapter to be an exhaustive account of what has already been written. Rather than an authoritative account of authority, the previous paragraphs provide some important initial insights into its nature and dynamics. Authority broadly refers to the capacity to produce communications that are willingly accepted by others out of trust and respect or, as Lincoln (1994) has emphasised, the strategic willingness to act as if this were the case. But the term also refers to the actor who is perceived to have this capacity within a given situation. Looking across the previous sections, I take this capacity not so much to be a stable and static quality, possessed by certain actors. Rather than an entity or a thing that someone simply has or does not have, I have come to view authority a result of situated relational interactions or the "kinetic outcome of social activity" as Pruitt puts it (2011: 24).[63]

More precisely, to understand where authority comes from and how it functions one can discern three layers; (1) the relationship between authorities and their subjects; (2) the character of the situation; and (3) the cultural context which conditions expectations and ideas of authority. In short, authority is constructed in relations, situations and within cultures. Authority implies an asymmetrical relationship that people willingly acknowledge. This experienced asymmetry hinges on inequalities of access to knowledge or latent power and secures not just individuals' attention but also their trust, respect, and acceptance. The recognition of authority's asymmetry lies vested in situations: when the many givens of a setting predispose an individual to define a situation as an authority situation in which their compliance is called for. And the culturally shaped values, orientations and expectations of people are of importance for the responses to authority claims. Table 1 below summarises this idea and forms a starting point for the rest of this book.

So far, our focus has been primarily on human authority. Apart from some passing references, not much has been written about the authority of (digital) things. Perhaps it is assumed that the concept of authority is either *a priori* confined to the human realm or that it does not need much further discussion when used synonymously for the increasing influence of modern-day technologies. The aim of this book is to further explore and develop the idea of digital authority, particularly in relation to the field of security. What I am interested in, is whether it has begun to make sense in our digitizing world to speak of digital authority, and if so under what conditions. Moving from the previous discussion that helps us understand authority generally to the specific case of digital authority begins with reiterating the definition of "digital authority":

> Digital authority is the capacity of a digital technology 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. It also refers to the technology that is perceived to have such a capacity within a given situation.

Digital authority, like its counterpart human authority, is enacted through communicative conventions that shape situations and orient the actions of its subjects.

The following exploration of digital authority is based on the previous discussion and the framework developed here. By looking into the relationship between technologies and people, the situation they engage in and the cultural background against which their interactions take place I further investigate the extent to which digital technologies can be authorities.

3.7 CONCLUSION PART 1

The first part of this book discussed the context of digital authority. Why should we be talking about digital authority in the first place? I described the digitization of our world and the increased role of digital technologies in private and public decision-making contexts. Many decisions now result from the interactions between human decision-makers and digital technologies. Ideally, these technologies *support* cognitive processes and result in better decisions by storing and providing relevant information to a human decision-maker. However, it is found that digital output can easily become a heuristic replacement for comprehensive information seeking and evaluation: computer says no. This becomes problematic when the technology turns out to be wrong and fails to live up to people's expectations. Against this background, I turned to the concept of authority to understand people's blind reliance on digital technologies. Authority refers to the capacity to produce communications that are accepted out of trust, respect, or the willingness to act as if this were so. It is rooted in relations, situations and in cultures. Authority implies an asymmetrical relationship that is recognised within a given situation by a culturally conditioned subject. While much has already been said and written about authority, the authority of digital technologies has remained largely out of scope.

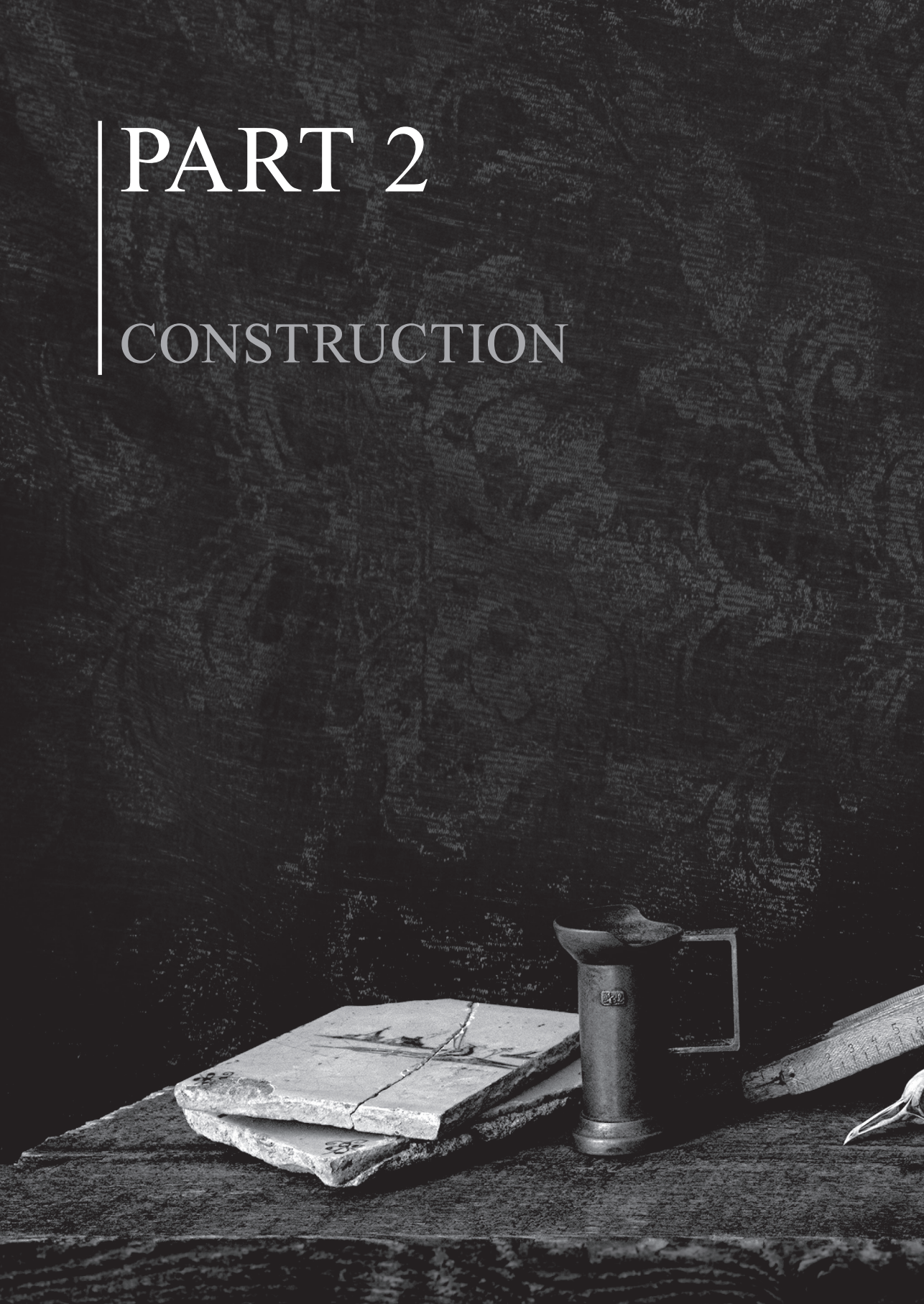
The next part of this book consists of three chapters and will look further into the idea of digital authority. Chapter 4 focuses on the *relationships* between technologies and humans. What does an authority relationship between a technology and human being look like? Subsequently, in chapter 5, I turn to the *situation* in which individuals may come to recognise a digital authority. In what environments would people recognise digital authority? Chapter 6 explores the *culture* of digital authority. Does today's cultural context facilitate digital authority?

Table 1. Layers of authority

	Relation	**Situation**	**Culture**
Characteristics	Experienced asymmetry	Defining an authority situation	Expectations of authority
	Consequential communications	Impressions	Conditioned values and beliefs
	Differentiated access to knowledge and force (persuasion and coercion)	Symbols (uniforms, titles, insignia, behaviors, places, etc.)	Dynamic and open to change

PART 2

CONSTRUCTION



4 RELATION

In 2015, two days before Christmas, there was a press conference in Washington. Governor Jay Inslee told the public that there had been "serious errors with serious implications". He revealed that over the course of thirteen years, thousands of inmates were mistakenly released early from Washington state prisons because of a "computational issue" (Washington governor, 2015). The problem was discovered by the family of an assault victim who had noticed that the perpetrator was about to get out of prison before the date the court had ordered. After they had contacted the Department of Corrections, it was discovered that for more than a decade, the department's software had been erroneously calculating the prison sentences of certain inmates.

It would take another three years before the public learnt about the situation. Meanwhile, more than 3,200 prisoners had been mistakenly released. In some cases, the consequences were serious. One of them, for example, committed an attempted robbery leading to the death of a 17-year-old. Another inmate became suspected of vehicular homicide and felony hit and run. High-ranking state officials had to apologise for the fact that this had happened while the suspects should have still been under lock and key.

The controversy in Washington caused widespread outrage and the agencies involved faced claims from the victim's families as well as from prematurely released offenders themselves who were taken back into custody years later to serve the remaining time (AP, 2018). In front of a state Senate's committee hearing, Dan Pacholke, the Secretary of the Washington Department of Corrections, was asked why the computer's sentence calculations were never checked by human officials. "I don't believe there's a performance system in place that is going to catch a mistake in a computer algorithm", he responded. The chair of the committee commented that he refused to accept this explanation and stated: "this wasn't a computer mistake. This was a people mistake" (Padden, 2016). Pacholke stepped down in the weeks after the hearing. In his letter of resignation, he again apologised for the "tragic consequences of this error" (Pacholke, 2016). At that time, Governor Inslee had already announced a software fix for the sentencing calculations. Until then, however, he insisted that all the upcoming release dates were to be calculated by hand.[64]

This painful story reveals people's willingness to rely on the communications issued by a computer. For years, the computer's output directed the actions of the personnel in Washington State prisons. In the previous chapter it was already stated that the literature on authority almost exclusively focuses on the relations *between people*. This chapter is about the relationship between humans and technologies. Can authority emerge here, and if so, how? Paragraph 4.1 discusses the work of North American philosopher of technology Don Ihde to understand the different ways individuals can relate to technology. In paragraph 4.2, I argue that digital authority represents a newly emerged relationship between humans and technology. I build on this observation and describe the capacity to communicate as a necessary condition for authority. In the previous chapter I concluded that authority always involves acts of communication. To be consequential, technologies thus need to communicate, and this ability is further discussed in paragraph 4.3. The subsequent paragraphs focus on the asymmetry characteristic of authority relationships and the potentialities of persuasion and coercion as constitutive parts of this. Paragraph 4.4 introduces the case of parking enforcement to illustrate the posited asymmetry between digital technology and its users. Paragraph 4.5 discusses how the implementation of digital technologies can severely limit individuals' opportunities for knowledge acquisition. In paragraph 4.6, in turn, I describe how digital authority depends on presumptions of differentiated access and the latent presence of persuasion in the background of the human-technology relation. Paragraph 4.7 considers the potentiality of coercion as a constitutive part of digital authority.

4.1 RELATING TO TECHNOLOGY

Intuitively, digital technologies and other technological artifacts are passive tools, not authorities. They are there to be used and to fulfill specific functions. Philosophers of technology, however, have already compellingly highlighted that technological artifacts cannot be well understood when approached as mere means or instruments to achieve our goals. Instead, the North American Philosopher of Science and Technology Don Ihde (1990) developed a (post)phenomenological understanding of human-technology *relations*, making visible how technologies co-shape human experience and action in the world. In his major work *Technology and the Lifeworld*, he distinguishes four different relationships human beings can have with technological artifacts. First, our involvement with the world can be *mediated* by a technological artifact (Ihde, 1990: 72-97). Ihde describes two forms of this technological mediation. Some technological artifacts merge with our bodies and from this position influence the way we experience our surroundings. Following Maurice Merleau-Ponty, Ihde calls this the *embodiment* relation. Examples of artifacts that form this kind of relation with humans are hearing aids, glasses, canes, and telescopes. When in use, these artifacts withdraw from our attention - they become perceptually transparent

and the world is perceived *through* them, accordingly (Ihde, 1990: 72, 86). Ihde schematises embodiment relations as:

Embodiment relations: (Human – technology) → World

In the other structure of mediation, the *hermeneutic* relation*,* an artifact is also mediating our perception, but the artifact is non-transparent (Verbeek, 2005: 126). Rather than simply withdrawing from our experience, the artifact provides us with a representation of the world that requires our interpretation. The term hermeneutics comes from the Greek verb *hermeneuein,* meaning "to interpret". Ihde uses the thermometer as an example. When you are in your home and you read the thermometer that is nailed to a fence outside, you *hermeneutically* know how hot or cold it is (Ihde, 1990: 84-85). In such situations, we experience reality *with* technology. Ihde schematises hermeneutic relations as:

Hermeneutic relations: Human → (technology - World)

A second set of human-technology relations Ihde identifies is *alterity relations*. Contrary to the previous relations, in alterity relations human beings are consciously involved with a technological artifact. It is the experience of a direct relation *to* a technological artifact (Verbeek, 2005: 126-7). The term alterity is rooted in the work of the French philosopher Emmanuel Levinas who refers to it as "the radical difference posed to any human by another human, an *other*" (c.f. Ihde, 1990: 98). This experience of *otherness*, Ihde says, we can also have with technological artifacts. Alterity relations can emerge for example when people use an ATM, encounter a robot, or prompt a chatbot - and with many other digital technologies. Ihde explains that technologies are experienced as others in alterity relations because they are seen to possess some sort of independence as well as give rise to an interaction between people and technologies (Ihde, 1990: 97 – 108). In alterity relations, as he puts it, there is always a "sense of interacting with something other than me (..), there is a kind of dialogue or exchange" with a technology (1990: 100). We engage in alterity relations with technologies because they appear to be autonomous – these technologies seem to have a life of their own (ibid.). As Peter-Paul Verbeek describes elsewhere:

> "Many toys, such as tops and music boxes, are fascinating precisely because of the apparent autonomy that they possess. Robots possess such an autonomy, to the extent that one could truly speak of "interacting" with these technological "beings". And today, automatic train ticket machines can not only take money and dispense tickets, but also give advice, provide route information, answer questions, and protest when something is done incorrectly" (Verbeek, 2005: 131).

Ihde schematises alterity relations as:

Alterity relations: Human → technology- (-World)

At the same time, Ihde observes that technological artifacts are, at best, *quasi-others*. He describes such otherness as "stronger than mere objectness but weaker than the otherness found within the animal kingdom or the human one" (1990: 100). Ihde compares the "spirited sports car" and the "spirited horse". Both can be used as an instrument to travel, but the horse is capable of resistance and disobedience, whereas a car can only malfunction. A horse can also live on its own, while the car needs to be started to be "animated". Moreover, the car will not be startled when a rabbit suddenly crosses the road, just like most horses will not simply follow the "command" to run into a wall when the driver is too drunk to notice (Ihde, 1990: 99). The animation we ascribe to a car, like any other technology, is of a different sort than that of a horse (Verbeek, 2005: 127).

The last set of human-technology relation Ihde describes is the background relation. Here, technological artifacts do not play a central role in our experience or interactions with the world, but reside in the background of it (Ihde, 1990: 111). The heating and cooling by automated thermostats, the light of lamps and the zooming of our computers are all examples of technologies with which we form background relations. While their effects may be subtle and often go unnoticed, their role is to influence how we experience the world. Ihde schematises background relations as:

Background relations: Human (technology/World)

4.2 AUTHORITY RELATIONS WITH TECHNOLOGIES

Verbeek (2008; 2015) has already observed that many recent technologies do not fit into Ihde's categories. For instance, a brain implant that is used for deep brain stimulation to treat certain diseases or psychiatric disorders, is more than embodied; it merges with the human body into a new, hybrid being. Verbeek calls this a cyborg relation: human/technology → World (Verbeek, 2015: 29). Another example are wearable technologies such as smart glasses that can be embodied to give an experience of the world, while also giving a representation of the world in a parallel screen. Verbeek speaks of this combined embodiment and hermeneutic relation as *augmentation*:

(Human-technology) → World + Human → (technology – World).

Also, authority relations with technologies do not fit Ihde's human-technology relationships. On the one hand, technological authorities play a mediating role in the relation between

humans and their world by actively shaping perceptions and actions. Such technologies tell people to think or do something by presenting them with a set of instructions or representations of reality ("your destination is on your right", "release inmate", "this is a risky individual", etc.), resembling a hermeneutic form of mediation that shapes experiences and interpretations which inform behavior. But with authority, on the other hand, the technology becomes the terminus of our experience in similar ways as with alterity relations. To recognise authority means to be directed at another actor that tells you what to think or do. This implies that not all technologies are potential authorities. A pair of glasses, a table or a broom will not trigger people to categorise them as authorities as they are not a source of interaction; but with digital technologies we do experience a sense of otherness based on their apparent autonomy and the interactions they give rise to. Yet, authority relations with technologies go beyond simple alterity as well. As discussed in chapter 3, it is the authoritative identity of an actor that motivates people to accept in advance whatever is communicated. In the authority relation, the influence that technologies exert on human perceptions and behaviors is so compelling that it becomes characteristic for the relationship between human beings, technologies, and their world.

As such, the digital authority relation is not so much about our interpretations and behaviors in the world *with* or *to* a technology but *by* or *according to* a technology. This could be schematically indicated as[65]:

Human → technology → human → World

4.3 WHAT THINGS SAY

It has already been stated several times that authority is considered an essentially social concept. But what makes this so? What is it about authority that would confine it to the realm of humans? It is in our use of language, or put more broadly, in the way we communicate. In paragraph 3.2, I discussed the work of Bruce Lincoln (1994) who linked authority relations to the capacity to produce consequential communications. Authority is an act of communication that incites accepting and compliant attitudes of an audience. And since such communications could always only be issued by people, authority remained tied to human relationships.

This is not to say that things did not say things. Scholars have also highlighted that many technologies send out messages on action possibilities. In paragraph 1.4, the notion of *script* was already introduced, referring to the prescriptions in technologies that steer users in how to relate, adapt and interpret a technology (Latour, 1992; Akrich, 1992). Another, well-known concept is that of *affordance*. Donald Norman (2013) describes an affordance as a design feature of an object that suggests how the object could possibly be used. Many of his examples concern buttons, doors and switches that are cognitively

associated with possible actions. And in the case problems arise – people that push the wrong side of a door or when automatic doors open unexpectedly – it is not necessarily the user's fault but rather the technology that was designed wrong and conveyed the wrong signal, Norman says (2013). That is, an affordance that would direct the users towards suitable behavior apparently is missing. Furthermore, Thaler and Sunstein (2008) have put forward the concept of *nudge* for situations were technologies make subtle changes in the way choices and information are presented with the aim of guiding people towards desired choices and behaviors. Defaults are a good example of how individuals might be nudged by digital technologies into certain behaviors – think of tipping defaults and privacy settings.

Still, these are not the communications emblematic of authority. The messages of these technologies that guide and influence human behavior are implicit; they only communicate in a metaphorical sense. They are prescriptions encoded into the technology, inviting, or stimulating behaviors without making apparent claims. Notwithstanding the "affordances" of analogue technologies that amplify or reduce aspects of experience or the scripts that actively invite certain actions while discouraging others, such technologies do not explicitly command or advice (Van de Voort, Pieters and Consoli, 2015:42-4). For this reason, we should not speak of the authority of, let us say, hammers, canes, or coffee cups - they did not issue communications to be believed or followed. While the script of a speed bump can have a similar effect as the words of a police officer, we do not recognise it as an authority. The communications of a speed bump are implicit, and we tend not to focus on it too directly. And when subjected to nudges, we are often unaware of the influence that is exerted on us in the first place. With authority, however, the communications to be believed or followed are actively attributed to the other – in this case to the technology that communicates something and, as such, becomes a reason for submitting to it.

One question that arises here is whether we are really directed at the technologies in the first place. Do we recognise the communications coming from a technology, or is it their makers who we are hearing? Doing what a technology tells us can also highlight the authority of the designers using other means, rather than that of the technology. After all, technologies are designed and (still) lack *intentionality* (Searle, 1980) – no matter how intelligent they may act, they do not "know" what they are doing. While stop signs and traffic lights tells us to stop, we may decide to do so not because of the technology itself, but out of trust and respect for those who made them and the possible consequences for non-compliance. But the emergence of digital technologies potentially challenges this. This claim is not necessarily grounded in the extreme cases of Artificial Intelligence that make headlines, in which individuals have seen the "ghost in the machine" (The Washington Post, 2022). Such cases are about whether AI, now or in in the future, can develop consciousness. Surely, when technologies would, for example, indeed become conscious or truly emotionally intelligent – having or exhibiting innate emotional capabilities (see Picard, 1997) – one could argue that it is the thing itself that has become

the one communicating as it has detached itself, like a child from its parents, from any human designer. Instead, building on the previous paragraph, I argue that we attribute all sorts of communications to less spectacular digital technologies, too. When interacting with many digital technologies, we tend to respond to their sounds, spoken words or written texts as if it were communications of their own – regardless of the human programmer lingering in the background. It becomes the technology that communicates to us. This is discussed in more detail in the next chapter.

Let us conclude, for now, that many human-technology relations are now characterised by digital communications that are actively responded to by human beings. I now continue the discussion on authority relations and its characteristics. In the previous chapter, the authority relationship was characterised as unequal, or fundamentally asymmetrical which produces attitudes of trust, respect, and acceptance. The rest of this chapter is dedicated to better understanding the potential nature and origins of this asymmetry when thinking about digital technologies. Let's start this off with a discussion on parking.

4.4 THE EVOLUTION OF PARKING ENFORCEMENT

Parking used to mean exactly what it sounds like: creating a park. In the spring of 1870, a law was passed by the United States Congress that authorised the city of Washington D.C. to reserve up to 50 percent of a street's width for the creation of "parks for trees and walks" (Richmond, 2015: 20). In the years that followed more than 70,0000 trees were planted on the side of the streets of Washington. The first parking system that followed thus had nothing to do with cars but was as Richmond observes "an early street tree system where parking defined the planning of trees, gras, and flowers along the side of roadways and the creation of sidewalks for pedestrians" (ibid.).

But it did not take long before the emerging roadside parks were used for more than just walking. People started tying up their horses and carriages to the parking trees when they were out in stores or visiting friends. This became the new "parking" and while it quickly became illegal to do so because it would hinder traffic and injure the trees, many took the risk of getting fined because of the convenience and shade that the trees offered for the waiting horses and carriages.

In the twentieth century cars came to dominate the streets. In the United States, the number of motor vehicles increased from approximately 8,000 in 1900 to more than 26 million in 1930 (FHWA, 2024). And just as people had tied their horses to a tree before, they now stopped their cars next to the parking strips on the side of the streets. City officials started cutting down trees and widening streets to deal with the sheer number of cars, thereby officially replacing the initial meaning of parking as a place where trees and bushes were planted with parking as a place to leave a car (Richmond, 2015: 21).

As car ownership increased, on-street parking quickly became scarce. Drivers could no longer park whenever and wherever they wanted, leading them to cruise around looking for a vacant space, wasting time, fuel, and congesting traffic (Shoup, 2011:1). Many municipalities struggled with the disordered bunch of parked cars and the city of Los Angeles even at some point experimented with banning on-street parking completely during daytime hours (Marshall, 2014: 362).[66] Other cities also tried to regulate parking, for example with posted parking time limits, but it turned out that these were very difficult to monitor. After all, officers could only observe and memorise the parking behavior of a limited number of vehicles within a certain area.

The introduction of a new technology in the 1930s, the parking meter, offered a solution as it allowed for keeping track of the time a vehicle was parked in a specific place. After inserting coins into the device, someone could park his or her car in the spot for a certain amount of time before more money must be inserted. When the person failed to do so before the time ran out, he or she could get a fine from observant officers checking the meters.[67]

Over time, most coin-operated, single-space, parking meters were replaced by electronic devices which covered many spaces within a defined area (Ison & Budd, 2016: 151). Initially, people needed to indicate on an electronic screen for how long they were planning to use the space, pay by cash or card and then received a printed ticket to be displayed in their vehicle. With newer systems, however, a printed ticket is no longer required. Modern parking machines and parking applications on mobile phones only require drivers to enter their vehicles' license plate. Parking officers, in turn, walk the streets and use handheld computers to check the plates of the vehicles parked in a particular area to ensure the parking session is valid.

Meanwhile, the next step in the evolution of parking has announced itself: scan cars. Scan cars are equipped with cameras on the roof that, while driving, read the license plates of parked vehicles to detect parking violations. The system compares the scanned plates to a database of paid parking transactions. Subsequently, in the case of complete digital enforcement, fines are sent automatically to the homes of the owners of the vehicles that are not found in the database. In many cities, however, human officers are (still) the ones deciding about the follow-up. The technology's alert of an illegally parked vehicle is then sent to a desk-officer who checks the scans and decides to issue a parking ticket or to dispatch a parking enforcement officer on a scooter to evaluate the situation on the street – for example to make sure that someone was not buying a ticket at the time the car was being scanned or to check for emergencies and other exceptional circumstances.[68] When the officer, in turn, determines that a car is, in fact, parked illegally, a fine will be issued via the system and sent to the registered owner.

A scan car reads more than 1,000 license plates per hour. This has not only lowered enforcement costs but also tremendously boosted city's incomes: digitisation resulted in an

increase in parking fines. Moreover, the use of scan cars influenced motorists' behavior: as the chances of being caught increased, more and more motorists decided to pay for their parking or stopped parking in paid areas. It is therefore no surprise more and more cities throughout the world are adopting scan cars for parking enforcement.

Technological innovation has been central in the evolution of parking as well in the work of those enforcing parking regulations. When parking spaces became a scarcer commodity and regulation became a necessary evil, officers became responsible for ensuring that drivers would comply with local parking laws and respond to violations. This initially meant relying on human senses and memory. Later, parking meters supplemented officers' observations. Coin-operated meters and printed tickets placed behind vehicles' windscreens communicated (non-)compliance, allowing for more effective modes of patrolling. Consequently, parking enforcement officers became constrained in acquiring information themselves as they now covered larger areas, making it increasingly difficult for them to monitor every car. With the introduction of digital systems where drivers register their license plate and sign up for a virtual parking ticket, information resources available to officers got limited further. For determining parking (non-)compliance, enforcement officers now needed to rely on that what is presented to them on the screen of their handheld computer. While they might, coincidentally, have seen someone buy a ticket themselves allowing them to assess the adequacy of what is in the system, most of the time they are unable to do so as they are patrolling large areas and physical evidence (e.g., a printed ticket or a meter on the sidewalk) has been designed away. With the introduction of scan cars, the relationship between enforcement officers and the availability of information in their work-environment has become even further obscured. Officers are no longer found wandering the streets but are only "on the spot" after they have been alerted by the digital technology. Their work is now highly structured around the communications of a digital technology. They no longer have a general overview of parking in the city themselves, but only through their computer. For the modern parking officer, additional information can only be acquired after they arrive at the place where the system has sent them.

4.5 THE ONES TO SPEAK IN AN EPISTEMIC NICHE

The story of parking makes visible how the relationship with digital technologies has become increasingly central in people's interactions with the world. With the introduction of digital technologies, we frequently see how the structure and sources of knowledge available to human decision-makers get severely constrained. Philosopher Jeroen van der Hoven (1998:100) speaks of *epistemic niches*. Epistemic niches are characterised by a structural knowledge-scarcity, making it difficult to get around an information system: users must decide, often in short timeframes and are constrained in acquiring relevant insights from outside the niche. An air traffic controller in a tower, for example, has as

his most important source of information that what is presented to him on a screen (e.g., flight paths, altitudes, airspeeds, etc.). Similarly, modern-day parking enforcement officers are primarily informed about vehicles' valid parking tickets by a handheld computer. For them the possibility to make observations themselves and to acquire useful additional information is severely limited. In the case of the air traffic controller human senses just do not allow for it in the first place - they generally cannot see where planes are flying, at what altitude and how fast they are going. Digital technologies have become indispensable for their actions and decisions.[69] In other situations, relevant information can be acquired elsewhere, but the structure and nature of people's work may prevent them from doing so. While users of digital technologies can occasionally call and check with a colleague or do an algorithms' calculations manually, – as was done in Washington State prisons after the software mistake responsible for the wrongful release of thousands of inmates was discovered - such strategies do not seem to be feasible in the long run as they take up a lot of time and are not flawless either. Organisational rules and structures can make alternative methods of obtaining information not a very attractive alternative. Leaving a digital technology as pretty much the only ones to turn too.

Epistemic niches tend to promote relations of *epistemic dependence*. This notion was developed by John Hardwig (1985) and refers to the situation where people come to rely on the knowledge of others but cannot assess its truth themselves. For them the belief that someone (or something) else has good reasons to believe something is true, forms, in itself, a good reason to believe in it.[70] For example, we depend on physicists with respect to the truth of Einstein's theory of relativity or on medical specialists for the truth of their diagnoses (Hardwig, 1985: 339). Most of us lack the training and ability that is required to assess whether relativity theory is correct or that allows us to make our own medical diagnoses. Rather, we believe that Einstein or a doctor has good reasons for believing in their claims, and therefore believe in their claims too. In a similar way, users can come to epistemically depend on a digital technology: they cannot assess the truth of what the digital system they are working with claims to be the case. People using digital technologies face the problem of opacity and often cannot check or verify the system's communications (Hayes et al., 2020: 3). Very often users have no idea based on what an algorithm advises them. Still, they might have good reasons to accept what the system says – for instance because it is known to be reliable or when there is simply no other information available to them.[71]

In short, the use of digital technologies may come with an artificial knowledge vacuum: an epistemic niche. Here, the system presents its output as accurate and worthy of belief, while at the same time the user becomes constrained in acquiring knowledge from alternative sources (Van den Hoven, 1998: 103). With limited alternative sources of information available to those in an epistemic niche, individuals may come to depend on a digital technology, with little or no control over the accuracy of the information provided. They become epistemically dependent.[72] The outcome is an asymmetry between

the technology and the individual; with limited alternative sources of information available to us, our relationship to digital technologies comes to the front of our experience. Here, computers seem to have become pretty much the only ones to speak. In an epistemic niche, the way a digital technology presents the world, will be pretty much how the world will come to appear to its audience. As such, when decisions must be made and a technology convincingly presents its output, it becomes very likely that people will follow it.[73] It might even be tempting to say that in such environments digital technologies have authority by default – but this would be premature conclusion as I show in chapter 7.[74]

4.6 PERSUASION IN TECHNOLOGY

Oftentimes, however, there is no pure epistemic niche shaping the relationship between humans and digital technologies.[75] The rivers and lakes in which people drive their cars are visible to everyone. And a security officer can deduce that a 2-year-old is probably not a terrorist (see next chapter). What, then, can we find, in the human-technology relation, that would explain people's reliance on digital communications rather than on their own intuitions or the ideas of others? The remainder of this chapter elaborates more on the possibilities of persuasion and coercion in human-technology relations as constitutive of the asymmetry characteristic of authority relations. I start with the former.

Persuasion is an age-old phenomenon that in the last decades has also become linked to technology. In his book, *Persuasive Technology* (2003), Fogg suggests that technological artifacts can persuade people, just like discursive arguments, to change their attitudes and behavior. To change what people think and do, technologies can, for instance, reward or remind them of (un)desirable behaviors. Think of commercial websites that provide tailored suggestions about books you might want to buy or the screen next to the road that shows a sad face when you're driving too fast. Fogg also mentions a system which monitors handwashing in restrooms and sends a signal in case people forget (Fogg, 2003: 47). With the development of this strand of research digital technologies are now increasingly recognised as potentially persuasive actors. In fact, Fogg writes that computers have several advantages over human persuaders (Fogg, 2003: 7). Among other things, computers can be more persistent and do not get tired, they can draw on a vast storehouse of information and they may be ubiquitous – having the ability to be almost everywhere (ibid, 7-11). According to Fogg, these characteristics undoubtedly increase the persuasive powers of a technology.

But authorities do not need to persuade. Persuasion only is needed when there is no authority. Yet, the possibility of persuasion, when present in the background, is fundamental for constructing such relationships. As argued in chapter 3, the exercise of authority may hinge on the presumption that persuasive argument can and will be given when asked for. The work on persuasive technology, and our own encounters with such

technologies, highlights how persuasion can be considered as potentiality within human-technology relations. Especially the general ability of computers to access and manipulate large volumes of information is of significance here. People often presume that there is a world of data and complex algorithms underlying the digital output that is presented to them and that precisely the right fact, statistic, or reference from that volume of data can be brought forward to persuade (Fogg, 2003: 8-9). The computer is a *black box* to them. Users themselves may or may not be barred from these insights themselves. But it is the belief in the existence of good reasons for compliance and that these can be provided when asked to do so – a capacity for reasoned elaboration – that underlies people's compliant responses. The inequality that underlies digital authority, then, is grounded in the perception that one (the technology) is has superior access to knowledge than the subject itself – the experience that "some are wise and some are otherwise".[76]

Still, the actual persuasive powers of digital technologies can be rather static. Often a computer will not provide users with exact explanations for their output because it is not programmed to respond as such. When people want an explanation, they generally must search for one themselves – for example by looking into the data or investigating the algorithm. Or people may not understand the digital explanation that is given – bringing us back to the idea of epistemic dependence as discussed above. Moreover, the elaboration that a digital technology provides us with often requires much interpretation. The beep, blinking light, or brief written text a technology might present us, often resonates because it triggers elaborative schemata already present within us – it incites us to tell ourselves that it is safer to wear seatbelts, that there is a world of evidence suggesting that taking regular breaks from typing is good for your health, that the system is right most of the time, etc.

And very often, the technological capacity to persuade is mediated – depending on the presence of others. Frequently it is other people that would provide us with reasons why we should do or think the thing the technology says. Programmers, designers, system managers, implementation officers and the like tend to function as henchmen of digital authority and are thought to be capable of providing further explanation upon request. Again, not everyone will understand them, but it is more about the *belief* that logical evidence for digital claims exists somewhere. This idea can motivate people to follow the technology regardless. When users assume that good reasons exist for whatever they are told, they end up being persuaded in advance. As Lincoln concluded, authority, then, becomes a "time-saving device or shorthand version of persuasion" (Lincoln, 1994: 5).

4.7 AUTHORISATION AND COERCION IN TECHNOLOGY

Digital technologies do not hold official positions with formalised powers. Unlike police officers, schoolteachers, managers, or military officials they do not have formal capacities that could enforce behavior. Computers cannot fire, arrest, or ban people. In fact, the General Data Protection Regulation (GDPR) does not allow for automated individual decision making at all. As stated in article 22 (1):

> "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

So far, there is no law that grants digital technologies the legal right to rule.[77] However, institutions do involve them in many decision-making processes and grant them positions of influence. As discussed in more detail in the next chapter, digital technologies are now advanced to various formal roles that are typically associated with authority. In these roles, they are integrated into institutional networks that foster trust and respect for those who are part of them. The asymmetry between the technology and human subjects, then, is institutionally ascribed and the result of a process of *authorisation*. Here, digital technologies have gained the confidence or support of an institution or its relevant members (such as managers or leaders) who confer authority upon the technology. This can take similar forms as with human authority. For example, a store manager can decide to hire a store security guard to prevent theft, or a coach can select a team captain for leadership on the field. When authorised by respected others, they are able to attract attention and evoke compliance from customers or players. Similarly, an organisation can confer authority on an expert system by making it mandatory for staff to consult it when making decisions. While someone might still want to refute the output of the technology, a narrative of institutional confidence can cultivate respectful and trusting attitudes towards the technology, highlighting its special position, and ultimately discouraging opposition.

Still, this perspective already highlights the dependence of digital authority on humans. While there have been some phenomenological explorations of human submission to digital leaders – specifically robots (Gladden, 2014) - people still hold positions that enable them to promote and restrict access to position of authority. This is also true when thinking about the potentiality for coercion in human-technology relationships. Here too, coercion is often conditional and depends again on the support of humans in positions of power. Only through the active or passive involvement of another authority figure, the potentiality of force becomes present in the background of the human-technology relation. Individuals interacting with the technology can come to presume that their resistance will be logged somewhere and that ignoring the digital output could get them into trouble.

But down the line, the capacity of coercive force rests with the backing of a human authority figure and when they withdraw support, the technologies' access to coercive force fades.

However, there also appear to be digital technologies that can exert force against individuals who refuse to comply with instructions or demands more autonomously. This takes us to the domain of *techno-regulation*. In his famous book *Code 2.0*, Lawrence Lessig (2006) observes how legal scholars tend to use legal rules to remedy pressing societal problems but that there are also many other regulatory forces that may shape human choices and behaviors such as market forces (e.g., taxation or price mechanisms) or social norms. In similar ways, Lessig argues, individuals' behaviors are shaped by architectures (Lessig, 2006: 123). Not only are we affected by architecture in the literal sense of the word – the buildings and structures that physically organise our environment and as such affect what we can and cannot do -, we are also influenced by architecture in a more figurative sense. Lessig says that software and hardware are forms of architecture that "constitute a set of constraints on how you can behave" (Lessig, 2006: 124). As put elsewhere, they are forces that steer, guide, and influence behaviors and actions of end users (Van den Berg & Keymolen, 2017: 191).

The idea that rules and norms can be implemented into technologies to regulate people's behaviors is known as *techno-regulation* (Brownsword, 2004; Koops, 2011; Hildebrandt, 2011; Leenes, 2011; Van den Berg & Leenes, 2013; Van den Berg & Keymolen, 2017). The design of software and hardware can be used to deliberately influence the actions that are available to users. They can only use the technology in certain ways, predetermined by the designers or regulators. Any option of non-confirming behavior is designed out - maybe except for those higher up in the hierarchy - or non-compliance is made utterly unattractive. In essence, techno-regulation binds users through norms that are embedded in the technology thereby aiming to prevent or discourage undesired or illegal behavior. This also makes that techno-regulation is effective; individuals may attempt to resist compliance, but they will ultimately be compelled to adhere regardless.

An often-used example of techno-regulation is that of speed bumps (Latour, 1992; Brownsword & Yeung, 2008; Leenes, 2011). While there are different means to regulate traffic speeds, such as traffic signs, speed bumps are effective because they do so in a direct and binding way: they leave little room for drivers to divert from a regulator's intended course of action (Van den Berg, 2011:405). Surely, a driver may choose to be "disobedient" and drive over the speed bump at a high speed, but this is uncomfortable and may damage the car – unlike the traffic sign which can more easily be ignored. As a result, with speed bumps drivers are more likely to adhere to the traffic rules than when a traffic sign is posted (ibid.).

In a similar vein, techno-regulation has become closely associated with digital technologies. The computerised access gate can be seen as a digital equivalent of the speed bump. The gate only opens for people who are allowed to access, such as employees,

individuals with valid passports or tickets or those with appointments. For those not in the system, the gate remains closed. Some individuals might, in turn, try to gain access by climbing over or forcing their way through the gate, but this can be dangerous, painful, or physically impossible. Like a speed bump, a computerised gate effectively regulates the behavior of those encountering it.

Techno-regulation offers interesting insights into the potential for coercion in digital technologies. These powers are vested in the digital technology or system itself; no others are needed to execute them (Leenes, 2011: 147).[78] Humans are only needed in the initial stages to design and authorise the technology to regulates human behavior in specific ways. It reveals how some technological artifacts can secure compliance with the alternative of force present in the background of the interaction. Even when an individual disagrees, they may still choose to follow the instructions of the technology because they understand that failure to comply will result in coercion – the gate will not open, the form will not save, or the system would simply not respond. This potential for coercive force can lead individuals to unquestionably do as they are digitally told.[79] Especially with technology, there tends to be no room for negotiation or bargaining. The technology can and will use force if it has too - resistance will be met with similar responses every time. This also makes the possibility of coercion a potential element for the experienced asymmetry on which digital authority is built, securing the digital capacity to produce consequential communications. This, however, does not mean that digital authorities structurally resort to the use of force. Both human and digital authorities do not need to persuade or coerce. The moment it becomes necessary to rely on coercion or persuasion to get something done there is no longer authority but just influence through force or arguments. To be a constructive feature of digital authority, the potentiality of coercion embedded into digital technologies through techno-regulation must remain exactly that: a possibility latently present in the interaction between the technology and the user. The possibility that a technology will use force when confronted with disobedience may contribute to an individual's regard for the technology and the willingness to accept whatever it communicates.

4

5 SITUATION

It must have been a strange sight, to see a commercial airline pilot being stopped from entering his plane. However, the pilot's name, James Robinson, was on a government's terrorist "watch list" - an information system that has quietly expanded since the attacks of 11 September 2001. The system is used to screen airline passengers. Those whose names are on the list are displayed by a computer as "high risk" which can lead to further questioning and being prohibited from boarding an aircraft (Florence, 2006). While this technology makes sense from a security perspective and is designed to counter terrorism, it is also heavily criticised as prone to false positives and being applied too rigidly by its users. This was also the case with the pilot of the commercial aircraft, who was certified to carry a gun into a cockpit but could not access his own plane (CNN, 2008).

He certainly is not the only one who got into trouble. In 2008, *The Washington Times*, for example, reported that for years false computer identifications have prohibited Federal Air Marshals from boarding flights they were assigned to protect. These Marshals had similar names as terrorism suspects and mistakenly appeared on a computer screen as "no fly". "In some cases, planes have departed without any coverage because the airline employees were adamant, they would not fly (..). I've seen guys actually being denied boarding", one Marshall said in an interview (The Washington Times, 2008).

Individual citizens, too, get mixed-up for terrorists by the digital system on a regular basis. Former pop singer Cat Stevens, for example, was denied entry to the United States in 2004 after his name was found on the no-fly list.[80] There has also been substantial publicity for the case of a Malaysian Stanford University graduate student who was arrested at San Francisco-Oakland International Airport in 2005. In a public trial the government revealed that she mistakenly ended up on the list because an FBI agent checked the wrong box (Kravets, 2014). A federal appeals court ruled in 2019 that the government must compensate for the millions of dollars she had to spend on her case. And there are numerous other examples of people who were treated wrongly on the basis of the computer's output. Most illustrative of the system's authority are those in which children are central. For instance, in Florida, an 18-month-old girl was removed from a plane because she was flagged a terrorist (CNN, 2012). Other children endured frequent intensive frisks and questioning for the same reason.[81]

What makes security professionals behave like this - even when in some cases, as some admitted, this runs counter to their own ideas?[82] The concept of authority helps us to analyse those instances in which individuals rely on a digital technology's output rather than their own judgement, even though this might lead them to do things that appear rather odd or downright wrong. To further explore this, I now move from authority relationships with digital technologies to authority situations.

Situations have always been important in the study of authority. In the beginning of book, Milgram's obedience experiments were already briefly introduced. Milgram claimed that people's extreme willingness to recognise and obey authorities should not be attributed to a particular personality or character as was done by others (i.e., Fromm, 1936; Adorno, 1950) but rather lies vested in the situation. As he acknowledges (Milgram, 1974: 5-6), this idea is highly reminiscent of Hannah Arendt's (1963) reports of the trail of Eichmann in Jerusalem. Arendt claimed that Eichmann, who was responsible for the logistics of deportations during the Holocaust, did not appear in court as a sadistic monster as the prosecution depicted him, but rather came closer to being a dull, ordinary man; a bureaucrat at a desk who simply followed orders as part of his job. For many, however, this appeared feigned innocence, and Arendt was heavily criticised for her notion of the *banality of evil*. It was felt that the atrocities of Eichmann surely required some sadistic, cruel, and perverted personality. But Milgram observed that his studies exactly affirmed Arendt's conception and the idea that ordinary people can go very far, even harming others under the sway of authority. The explanation for this he found not in a peculiarly aggressive individual personality but rather in their definition of the situation (Milgram, 1974: 138-139; Rood, 2013: 113; Blass, 1999: 958).

This chapter explores the notion that authority lies vested in the character of the situation. It considers how digital technologies can shape authority situations for us. Paragraph 5.1 first describes that people define the situations they find themselves in. Using the work of Erving Goffman, I argue how these definitions form the basis for how individuals present themselves and act in social settings. The following paragraph (5.2), explains that these definitions, in turn, are grounded in learnt scripts that guide our attention, comprehension and expectations of what is going on and how to respond appropriately. Paragraph 5.3 discusses how these scripts guide the ways people recognise and respond to authority situations. The right situational cues can easily trigger the accepting, reverent and submissive responses associated with authority. But how would this work for digital technologies? The subsequent paragraphs investigate the ways digital technologies mimic authority situations by tapping into pre-existing social scripts. Paragraph 5.4 and paragraph 5.5 present two strands of prominent research (anthropomorphism and the Media Equation) on Human-Computer interaction showing that people often respond to technologies in very similar ways as they do to other people. Both knowingly and unknowingly we use all sorts of social categories and expectations in situations where we interact with computers.

This discussion enables us to transcend the distinction between social and technological actors, fostering the notion that human beings can come to perceive digital technologies in similar ways as human authorities. Subsequently, in paragraph 5.6, I provide some more specific insights into the technological cues that could transform the interactions with computers into digital authority situations. In this final paragraph (5.7) of this chapter, in turn, I wonder if it is not just the designers, rather than the technologies, whose authority people might recognise here.

5.1 WE HAVE A SITUATION HERE!

In the 1920s, William Isaac Thomas already highlighted the importance of situations for social scientific research. He contended that for understanding human behaviors one must investigate the situations in which they occur (Thomas, 1923: 230-1). For Thomas, situations are not just the objective elements such as the physical characteristics of the environment or the economic or political situation in which the individual operates, but also have a more subjective side that is associated with how individuals perceive and give meaning to the situation (Thomas, 1923: 42). These subjective elements can lead people to respond to situations rather differently – grounded in a diversity of past experiences, cultural backgrounds, and ideological orientations.

Thomas thus contended that is not just the objective features of a situation people respond to, but also to the meaning the situation has for them (Merton, 1968: 475-6). He writes that "preliminary to any self-determined act of behavior there is always a stage of examination and deliberation which we may call the *definition of the situation"* (Thomas, 1923: 41). And once individuals have ascribed some meaning to the situation, it will shape their consequent behavior. But, as Thomas explains, "there is always a rivalry between the spontaneous definitions of the situation made by the member of an organised society and the definitions which his society has provided for him" (Thomas, 1923: 42-3). People are born into a culture, where they learn in social groups about basic values, norms and expected behaviors in social situations. But despite the fact that definitions of the situation are learnt in social groups, there is always some room for personal interpretation. Among other things, because in many situations cultural definitions remain vague or too abstract. Therefore, as concluded elsewhere, "definitions that ultimately determine conduct are a result of attending to certain features of a situation, and then resolving at times a major tussle, and at others a minor tussle, between cues provided by "society" and the spontaneous definitions made by the actor" (Perinbanayagam, 1974: 523).

The notion of the definition of the situation is a prominent feature in interactionalism and the work of Erving Goffman. We spend most of our life, Goffman says, in situations where we interact with others and must choose how to respond appropriately. This includes everyday environments in which most of our social interactions take place - such as schools,

work, birthdays or in restaurants - as well as the more unusual settings we only occasionally find ourselves in – think of weddings, funerals, or job interviews (Goffman, 1983: 2). When entering a new situation, individuals face the question: "what is going on here?" (Goffman, 1986: 8). People use the definition of the situation to know what is expected of them within that setting. They ascribe meaning to the general environment they have entered and provide a sense of the role of others present in it so that they know how to behave in a meaningful and socially accepted way. Based on their definition of the situation, people choose a certain role to play and behave in a particular manner. A definition of the situation thus generally refers to how our interpretation of the social setting we engage in shapes our expectations of what we must do, who we will interact with and with what aim.

Much of this, however, happens unknowingly. As Goffman writes, the use of a definition of the situation does not necessarily involve a conscious process of deliberation and we tend to only explicitly ask ourselves "what is going on here?", when actively confused or in doubt - in times things are unclear to us and we do not seem to immediately know the answer. But in moments of relative certainty, the question is asked only "tacitly" (Goffman, 1986: 8). As also explained by Van den Berg (2009):

> "The answer to the question becomes visible in the choice of a role and its accompanying actions, but emerges more often than not in such an instantaneous, automatic way that the individual won't even be aware of the fact that he or she is using a definition of the situation at all to assume a role within the given context. Rather, a definition of the situation generally is nothing other than the implicit process of interpreting and ascribing meaning to the setting one enters, a process so subtle and unthinking that agents are mostly unaware of it taking place." (Van den Berg, 2009: 190).

In short, the interactionist notion of the definition of the situation highlights that the roles human beings take up and the behaviors associated with them are situationally specific or conditional - depending on how they interpret the social setting they become part of.[83] In the next paragraph, as well as in chapter 6, I further emphasise that these definitions of the situations are, of course, thoroughly imbued with social norms and standards. For now, let's conclude that the concept of the definition of the situation provides useful insights about how people interpret and act in social situations. This forms an important starting point for understanding (digital) authority situations. But first let's look a bit more into the mechanisms behind our situational definitions and responses.

5.2 DEFINITIONS, CUES AND SCRIPTS

As social beings, we often simply "know" a social situation and what behaviors are called for there. The concept of *script* is valuable because it explains the (implicit and explicit) knowledge of people at play when this is the case. It has been used to study human behavior in a wide range of different fields (Albarracin et al., 2021: 1).[84] At their core, however, is often the interactionist idea that members of a social group share a common body of knowledge that prescribes situational modes of being that allow them to act in appropriate ways (Abelson, 1981) – like actors sharing a script that tells them what role to play.

Perhaps the most influential interpretation of scripts is by Roger Schank and Robert Abelson (1977) who introduced the concept to the field of cognitive psychology and Artificial Intelligence. In their book *Scripts, Plans, Goals and Understanding* (1977), Schank and Abelson studied the nature and use of knowledge to further the field of Artificial Intelligence. They wondered how humans make sense out of what they hear, see, or are told about the world (Schank & Abelson, 1977: 41). What knowledge structures are in place to understand and function in their world? How do individuals decide what to do and understand the roles of others? In other words, where do our situational definitions and actions come from?

Schank and Abelson coined the term *script* to refer to the generic knowledge structure that organises our knowledge about activities we are familiar with (Schank & Abelson, 1977: 208). Scripts can be defined as a "pre-determined, stereotyped sequence of actions that defines a well-known situation" (ibid). They are packages of knowledge in the human mind that we use to navigate everyday life. According to this idea, for every situation there is a sequence of conceptualisations, which are associated with the normal order of things. Hence, there is a birthday-party script, an office script, a classroom script, a supermarket script, a public transport script, and so on (Schank, 1975: 117). Each script specifies the roles of the actors and objects within a given situation and contains a series of actions. Consider the paradigmatic example of the restaurant script (Schank, 1975; 117-8; Schank & Abelson, 1977: 210-2; Abelson, 1981: 715-6). This script details, from the perspective of the customer, how events in a restaurant typically unfold. It involves several procedural steps such as "finding a table", "looking at the menu", "signal to waiter" and "ordering food".

The scripts stored in our memory allow us to quickly determine what is going on and respond in ways that fit that situation. This forms the basis for how we define situations and many of our routine behaviors. Scripts are activated automatically by the situation. The situational context of a restaurant, for example, activates a particular script that directs our attention and guides our behavior in that environment, while entering an office will trigger a different script, containing actions such as checking in with colleagues, getting coffee and sitting at a desk. When there is an evoking context, Abelson writes (1981: 791), individuals will enter a script. That is, when people recognise certain cues in the situation,

they are triggered to read and respond to a situation in a categorical manner. These may be signs present in the physical environment as well as in the actions of others. Scripts are thus anchored in a specific social setting and help individuals infer their role and how to proceed.

Scrips are important as they provide most of the tacit, or background world knowledge that guides our attention, comprehension, expectations, and reasoning that help individuals to find their way in so many everyday situations (Graesser & Whitten, 2001: 261). But where do scripts come from? They are learnt. We learn them throughout our lives. When we are young and learn so much about the world but also when we are older and old scripts get modified or new scripts develop when we enter unknown situations. As Schank (1995) explains:

> "We learn them [*scripts*] by practicing them over and over. We can learn them as children, by being taken to a restaurant many times and gradually learning each step of the restaurant script. We can learn them through expectation failure by seeing an aspect of a script fail to be true (chopsticks instead of forks, for example) and explaining the difference to ourselves, creating a new Oriental restaurant addendum to that portion of the everyday restaurant script. And, we can learn them as adults, by, for example, going on our first airplane ride, trying to understand its script and gradually modifying our understanding with each subsequent trip" (Schank, 1995).

Note that because scripts are learnt, they are inherently cultural – they arise and exist in shared cultural meanings. Moreover, individuals often have a different set of previous experiences. This explains how different people end up using different scripts to understand and respond to the same social situation.

Scripts can and do frequently fail. Every once in a while, we are confronted with physical and social obstacles – we fall asleep in public transport and miss our stop, we forget to bring our wallet to a restaurant, we miss a meeting due to endless traffic jams on Monday mornings. These unusual cases then become part of our memory which can later be accessed to deal with similar obstacles, or when these occur frequently enough, the anomalies become part of the typical script content (Graesser & Whitten, 2001: 261).

The use of scripts solves the problem of our limited information processing capacity (see also paragraph 2.3). Once learnt, and when no anomalies occur, no new conscious effort is required, allowing for routine response behavior. In fact, by drawing on the experiences from previous situations, no new behavioral patterns must be learnt for new, but relatively similar, situations (Bozinoff, 1982: 486-8). The restaurant script works in many kinds of restaurants and saves people the energy of deciding what is going on exactly and expected from them. This information is already stored in their memory. While people can always decide to behave differently and deviate from the appropriate script, often they follow

the familiar scripts that seem to reasonably fit the situation they are in. This is also what intelligence boils down to according to script theory: recognising and enacting situationally relevant scripts. And this is exactly where it becomes interesting when digital technologies enter the stage.

5.3 AUTHORITY SITUATIONS AS ROLE CUES

The notions of scripts and definitions of the situation are important for understanding how people come to recognise and respond to authority situations. In short, situations trigger scripts that help human beings to make sense of what is going on and react to them appropriately. And when the right authority cues are present in the situation, a script will be triggered that facilitates a definition of the situation and will lead individuals to take up the accepting, reverent or submissive roles that are so typically associated with authority relations.

Like with other scripted behavior, we learn to recognise and respond to authority situations. Much has already been written about how authority is a fact of social life, supported by widely shared norms in religious and moral teachings (Milgram, 1963; De George, 1985; Kelman & Hamilton, 1989; Cialdini, 2007). Cialdini, for example, refers to the Bible and describes how Adam and Eve failed to obey the ultimate authority, resulting in the loss of paradise (Cialdini, 2007: 216). In the Old Testament there is also the tale of Abraham and Isaac, revolving around the notion of respect and trust in divine authority. In this biblical story, Abraham was to sacrifice his only son because God, without any explanation, told him to do so. When Abraham was about to sink a knife into his son's body an angel stayed his hand and told him that God was only testing him and that he should not harm is son. In return for his unquestioning obedience, God rewarded Abraham by making his descendants a great nation. But also beyond the realm of religion people are taught that following authorities is the right thing to do. Take, for example, all the songs, stories and lessons revolving around respecting parents and trusting experts. From the moment we are born until the day they die; people are morally encouraged and socialised to rely on proper authorities.

There are many items, places and symbols that help us to recognise authority. In chapter 3 I mentioned both the tangible (uniforms, badges, diploma's, vehicles, special seats, locations, etc.) and intangible symbols (titles, language styles, demeanors, times, etc.) that mark the authority of those associated with it and help lay claim to people's attention, respect, and trust. One way to understand such symbols is as situational cues that guide people in how they understand and respond to the social setting they are in. Similarly, the observable behaviors of others – obedient, docile, or otherwise - also help to interpret "what is going on" and infer their role in an authority script. When recognised, all these cues offer a mental shortcut for identifying authority situations and choosing the appropriate responses: a script is activated, and unconscious or routine response

behavior becomes likely. About the actors that claim authority by actively utilising these authority cues, it could be said that they end up defining the situation for us. That is, their performance effectively turns a situation into an authority situation when the situational cues are recognised and responded to by others in that setting.

Milgram experimented extensively with these situational cues to explain the obedient responses of his subjects. For example, he investigated if subjects would be less compliant when given instructions by an ordinary man than with an authoritative experimenter (they were) (Milgram, 1974: 93-97; 99-105), whether the presence of rebellious peers mattered (it did) (ibid: 116-121), or what would happen when the subjects would be visible and audible to the subjects (obedience diminished) (ibid: 34 -55). In terms of scripts: Milgram was manipulating the situationally specific cues that would trigger authority scripts among his subjects.

Authority situations thus trigger scripts that inform individuals about what is going and what roles should be enacted. This way, *situations structure authority relations* and guide the accepting and compliant responses of individuals. But what does this all mean for digital authority? So far, it has been about situations where people interact with other people and enter an authority script. How would this, if at all, apply to digital technologies? These questions take us back to an earlier discussion on alterity and authority relations with technologies and the work of Ihde (1990). In the previous chapter it was argued that humans can relate to technologies as an *other* and how this forms an important condition for the construction of authority. Digitisation has generated new forms of direct interaction with technological artifacts. The second part of this chapter now focuses more specifically on how interactions with digital technologies, too, can generate authority situations for people via pre-existing scripts. It builds on the wealth of empirical research on Human-Computer Interaction (HCI) highlighting that situations with digital technologies can easily become social situations. That is, people tend to use all sorts of social categories and expectations in situations in which they interact with technological artifacts. Both knowingly and unknowingly, we treat digital technologies as if they were real others – following the same social rules in our interactions as we do with other people. Based on two strands of literature it can be concluded that people can read and respond to situations with digital technologies in similar ways as to situations with human authorities. The first strand of literature that is worth discussing in this regard is on anthropomorphism.

5.4 ANTHROPOMORPHISM: SEEING HUMAN

For a long time, anthropomorphism has been the typical explanation for social responses to nonhuman objects of any form. David Hume (1757: III) already spoke of the human tendency to conceive all beings like themselves. Similarly, Goethe once wrote that "humans never know how anthropomorphizing they are" (c.f. Guthrie, 1993: 63). Anthropomorphism is generally perceived as the attribution of distinctively human characteristics and traits to non-humans. When people anthropomorphise, they typically ascribe human threats to nonhumans. In the 6th century B.C, Xenophanes was the first one to speak of anthropomorphism to describe the striking physical resemblance of gods to their believers. In its current use, anthropomorphism goes beyond the simple attribution of superficial human characteristics - such as a humanlike face or body - to a nonhuman (Waytz, Heafner and Epley, 2014: 113). Instead, Waytz and colleagues define anthropomorphism as "the process of inductive inference whereby people attribute to nonhumans distinctively human characteristics, particularly the capacity for rational thought (agency) and conscious feeling (experience)" (ibid). And it is the possession of such higher mental states that are often the focus of both lay theories and philosophical definitions of personhood (Waytz, Cacioppo & Epley, 2010). As was concluded elsewhere, when anthropomorphised "nonhuman entities [become] social entities" (Caporael, 1986: 215) as they are attributed essentially human characteristics (i.e., a humanlike mind, capable of thinking and feeling).

In her empirical studies on anthropomorphism, Turkle (2005) discusses how children often think and talk about computers as a living thing. She writes that the interactivity and unpredictability of computers induces young children to attribute typical lifelike qualities to computers such as feelings, consciousness, and intentions. Turkle found that this tendency generally fades over time. Older children learn to make finer distinctions between computers and people (Turkle, 2005: 57). Once people cognitively develop and acquire more knowledge about computational objects, their anthropomorphic behavior often decreases (see also Winograd & Flores, 1986).

At least, that is what is generally the case. Sometimes, Turkle argues, even adults can be fooled by a machine into thinking that they are interacting with a human. She uses the well-known example of the ELIZA program that was composed in the 1970s by the computer scientist Joseph Weizenbaum. ELIZA was programmed to converse with people, playing the role of a Rogerian psychotherapist in an initial interview with a patient. Back then, before the development of AI driven language processing tools such as ChatGTP, it was still impossible for a computer program to engage in a conversation on a broad number of subjects. However, the Rogerian psychotherapist offered an opportunity to bypass this problem through a technique of reflecting to the statements of a patient. For example, when a patient said, "I need some help", the therapist would reply "I understand that you need some help". This technique made it possible for a "dumb" computer program such as

ELIZA to respond in a deceptively intelligent way - without actually understanding what is being said (Turkle, 2005: 42).

The ELIZA program became widely known for its impact on "patients", and this even perplexed Weizenbaum himself. In his book *Computer Power and Human Reason: From Judgement to Calculation* (1976: 6-7) he writes:

> "I was startled to see how quickly and how very deeply people conversing with [ELIZA] became emotionally involved with the computer and how unequivocally they antropomorphized it. Once my secretary, who had watched me work on the program for many months and therefore surely knew it to be merely a computer program, started conversing with it. After only a few interchanges with it, she asked me to leave the room. Another time, I suggested I might rig the system so that I could examine all conversation anyone had with it, say, overnight. I was promptly bombarded with accusations that what I proposed amounted to spying on people's most intimate thoughts; clear evidence that people were conversing with the computer as if it were a person who could be appropriately and usefully addressed in intimate terms. I knew of course that people from all sorts of emotional bonds to machines, for example, to musical instruments, motorcycles, and cars. And I knew from long experience that the strong emotional ties many programmers have to their computers are often formed after only short exposures to their machines. What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people. This insight led me to attach new importance to questions of the relationship between the individual and the computer, and hence to resolve to think about them."

Through ELIZA, Weizenbaum learnt about the computer's strong ability to convey the illusion of the possession of human characteristics and faculties. Even experienced computer users conversing with ELIZA were led to false beliefs about the expertise of the program and that it understood them as though it was another person, leading them to forget that, in fact, ELIZA was only a machine (Turkle, 2005: 42. See also Van den Berg, 2009: 252).

When a computer shows some degree of intelligence, it can become more than a machine for people, Turkle writes (2005: 247-8). In their interactions with computers, people then willingly ascribe human attributes such as feelings, intentions, and behaviors to it. She gives the example of an airline ticket agent (Jean) using a computer to make customer reservations:

> "Jean often confronts angry clients. The planes are overbooked; the computer has been programmed on the assumption that some people won't show, but occasionally everybody shows, and not everyone can go. Jean's excuse is, "The computer fouled up". I talk to her about what she means. Is this a "manner of speaking?" It is not. She seems to experience the computer as an autonomous entity that can act on its own behalf and is thus "blamable." (Turkle, 2005: 247).

While such anthropomorphism may be naïve, it is sometimes hard to resist, not just because of how computers are behaving but also because of their irreducibility, Turkle explains. For many computers are just black boxes. In Turkle's words, it is hard to "block the temptation to personify the computer by saying what it "really" is, hard to block the suggestion that the computer "thinks" by saying what it "really does"" (2005: 248). And since "computers are not "like" anything else in any simple sense", some people will think and speak of them as human, she concludes (ibid.).

In Turkle's research, people attribute human characteristics to nonhumans *consciously* (Van den Berg, 2009: 252). From this anthropomorphic perspective, it is not surprising that some people explicitly treat a machine as it were a real person – or, as I argue later, as an authority in the traditional, human sense. When there is a sincere belief that machines have (human) intentions and feelings, people will logically perceive them as warranting social treatment. Below I will further investigate what this means in relation to the situational layer of digital authority. Let us first focus on another prominent view on people approaching digital technologies as human beings. At Stanford University, Clifford Nass and colleagues conducted a series of empirical studies resulting in the so-called 'Media Equation' (1996). In this research it is argued that *all* people display social behaviors towards computers, even when they are very aware of the inappropriateness of such responses.

5.5 THE MEDIA EQUATION: COMPUTERS ARE SOCIAL ACTORS

In their book *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places* (1996), Byron Reeves and Clifford Nass argue that the interactions between humans and computers are fundamentally social. They go beyond the anthropomorphic idea that people's social responses to computers are the result of a social or psychological deficit and will go away with age, knowledge, and conscious thought (Reeves & Nass, 1996: 4). The Media Equation, in its most basic form, refers to notion that "media equals real life" (ibid: 5). Reeves and Nass sum up an overwhelming amount of experimental evidence showing that basically all human beings tend to treat computers and other technologies as social actors.[85] To elicit these behaviors, no fancy Artificial Intelligence or humanoid robots are needed. Even the most minimal social cues from a machine lead people to *automatically* expect them to follow social and natural rules – rules

that initially only applied to interpersonal interaction. Yet, most people are unaware of this tendency and are prone to deny the fact that they treat these machines as though they were another person. But *unconsciously* we do – all the time.

For example, one experiment found that people apply norms of politeness to computers. When a self-praising computer asked a user about its performance during a previous task, individuals turned out to be significantly more positive than when this evaluation was done on a different computer or in a paper questionnaire. Consistent with known human-human behaviors, we tend to be nicer to others in a face-to-face setting Nass, Moon and Carney concluded (1999). It appears that people do not want to hurt the "feelings" of a helpful computer.[86]

Another series of experiments focused on people's use of social categories and roles when interacting with a computer. For instance, Nass, Moon and Green (1997) demonstrated that people apply gender stereotypes when interacting with a computer. In line with earlier social-psychological findings, people perceive praise from male-voices computers as more friendly and valid than praise from female-voices computers. It was also found that participants regarded female-voices computers as better tutors on "feminine" topics such as love and relationships.[87]

These experiments, and there are many more, show us that social responses towards computers are not rare or unique. On the contrary, the treatment of computers as if they are humans is ubiquitous. It is commonplace across virtually all areas of social psychology and persistent among different types of people (Lee and Nass, 2010; Fogg, 1997; Nass and Reeves, 1996; Nass and Moon, 2000; Sundar and Nass, 2000). We all guide our interactions with computers by natural and social rules and respond as if we expect them to also follow these rules. Yet, social interactions with computers often manifest themselves in subtle ways. Rather than obvious, easy to observe responses such as yelling or talking to a computer, people's social behaviors are much more profound and they are often unaware of such behavior (Nass, Moon, Morkes, Kim & Fogg, 1997). But why do we respond to computers as if they were real people in the first place?

The explanation for the media equation is rooted in social and cognitive psychology, including the extensive work of Ellen Langer (1978; 1992; 2000; 2010; 2014) on mindful and mindless behavior. According to Langer mindfulness is best understood as a flexible cognitive state of openness to novelty in which the individual is occupied with drawing new distinctions and categories (Langer, 2000:220). People are then actively engaged with their environment, and open to noticing new things. However, social psychologists have found that many social interactions are performed mindlessly. When mindless, we rely on distinctions and categories drawn in the past and the risk is that we "remain blind to subtle changes in the situation" (Langer, E., Pirson, M., & Delizonna, L., 2010: 69; see also Langer, 2014: 289). That is, much of the available information in the environment is left unprocessed and the individual will operate from a perspective that is rigidly fixed.

Mindless behavior is rule- and routine-governed, largely independent of the immediate circumstances (Langer & Moldeveanu, 2000).

Mindless behavior is based on only a few contextual cues (Langer, 1992; 290; Langer, 2000: 220). As described above in paragraph 5.2, these cues easily trigger situational definitions and behavioral scripts that fix someone's attention to certain information and away from other - potentially contradicting - information (Langer, 1992: 299-300). In contrast to individuals who mindfully respond based on active processing of incoming information, the mindless individual prematurely commits him or herself to "overly simplistic scripts drawn in the past" (Nass & Moon, 2000: 83). Once the critical elements or cues appear, heuristics or cognitive shortcuts are activated; someone will act in a predetermined or automatic manner.[88]

Media equation research suggests that people's social responses to digital technologies are the result of mindlessness. When individuals mindlessly interact with these technologies, they tend to apply old, easily accessible, social scripts developed for human-human interaction. According to Nass and Moon (2000), this is because computers display certain cues that suggest humanness. These cues sway people to mindlessly categorise them as social actors and respond to them in corresponding ways (Nass & Moon, 2000: 83). At the same time, the contextual cues that expose the "essential asocial nature of a computer" - and thus the inappropriateness of applying social scripts in human-computer interaction - are not processed in a state of mindlessness (ibid; Sundar & Nass, 2000: 683). According to Reeves and Nass (1996) we are hardwired to socially respond to human-like cues from our environment. They say that these social responses are instinctive rather than rational and are deeply engraved in our evolutionary history:

> The human brain evolved in a world in which only humans exhibited rich social behaviors, and a world in which all perceived objects were real physical objects. Anything that seemed to be a real person or place was real. During all of the 200,000 years in which Homo sapiens have existed, anything that acted socially really was a person, and anything that appeared to move toward us was in fact doing just that. Because these were absolute truths through virtually all of human evolution, the social and physical world encouraged automatic responses that were, and still are, the present-day bases for negotiating life. Acceptance of what only seems to be real, even though at times inappropriate is automatic. Modern media now engage old brains. People cannot always overcome the powerful assumption that mediated presentations are actual people and objects. There is no switch in the brain that can be thrown to distinguish the real and mediated worlds. (..) Absent a significant warning that we've been fooled, our old brains hold sway, and we accept media as real people and places (1996: 12).

What is fascinating about this body of knowledge is that it reveals how little of a cue is needed to elicit social responses to computers. As Lee and Nass conclude "people are very liberal in assigning humanity to an artificial stimulus" (2003: 294). A computer system or interface does not need to look, sound, and move like a human to convey social presence. In fact, poor attempts to make a computing product as human as possible are likely to lead to experiences that are less human-like, uncomfortable, and less social (Nass & Gong, 1999).[89] In many of the experiments, computers only used written language and provided simple interactive options to the user rather than sophisticated features such as humanlike faces or voices. Most basic modern-day computing products are incorporated with such social cues – such as interactivity, the use of language (either in text or voice output) or they fulfill typical social roles - that can easily trigger people to mindlessly categorise them as human (see Nass & Steuer, 1993; Nass, Steuer, Henriksen & Dreyer, 1994; Moon & Nass, 1996; Nass & Moon, 2000).[90] Therefore, in many situations digital technologies will encourage social responses making their social treatment something ubiquitous (Nass & Moon, 2000: 84).

But although we struggle to respond to modern-day technologies in a manner that is consistent with their (non-living) ontological status, this does not mean that social responses to computers and other media are inevitable. Sometimes, the technologies themselves might suddenly remind us about their fundamental non-humanness (see also chapter 7). At other times, Reeves and Nass argue, we overcome our tendency to respond socially to a computing product through substantial cognitive effort and mindful reflection (1996: 13). For example, the moment when we explicitly tell ourselves that "it is just a thing" when we are on the verge of throwing our seemingly unwilling laptop out of the window. Such mental reminders offer us an escape from our primitive and mindless social responses. Yet, this is not what comes natural to us and is therefore difficult to maintain all the time, Reeves and Nass conclude. Let's now link these thoughts to the idea of authority.

5.6 TECHNOLOGIES IN AUTHORITY SITUATIONS

In the 18th century, the French physician and philosopher de la Mettrie (1747) wrote his famous, albeit controversial work *Man a Machine (L'homme machine)*. In his view, humans were little more than a complex and self-maintaining machine. He compared the human body to a finely designed clock: "The human body is a machine which winds its own springs" (1747: 5). Nowadays there are still many situations in which people think of themselves in mechanical terms. Especially popular in this regard is the use of a computational vocabulary. People may, for example, say they "cannot compute" when they cannot understand something or that they need to "reboot their life" when they do not like their current situation. As Turkle also writes, this computer language is not just a manner of speaking but "carries an implicit psychology that equates the process that takes place in people to those that take place in machines" (2005: 22).

The studies discussed above suggest that with the advent of the digital age, it is now also appropriate to reverse de la Mettrie's treatise and speak of *Machine a Man*. Modern-day technologies appear to us as increasingly autonomous and interactive which has made them a focal point of our attention. The result is that in today's world we do not only frequently speak of digital technologies in human terms, but we have also come to think of them and treat them as such as well. While many people might intuitively think of digital technologies as instrumental objects, most will inevitably interact with them as they would with other humans. By consciously anthropomorphising or through unconsciously equating media and real life, even instrumental objects become perceptual social actors.

This blurring distinction between humans and machines challenges us to reconsider the nature of many social phenomena, including authority. But what makes people respond to situations with digital technologies in similar ways to situations with human authorities, triggering their unquestioning acceptance of what is communicated? How does a technology generate an authority situation for people and trigger pre-existing scripts that will lead them to respond with trust, respect, and a strong willingness to do or think what they are told?

This brings us back to the discussion earlier on authority situations and scripts. The studies described above help us understand how cues from digital technologies can trigger social scripts and give rise to new authority situations. In contrast with many other inanimate objects, we now see digital technologies performing tasks and roles that were previously performed by human authority figures. We now interact with digital technologies in roles of doctors, teachers, experts, loan advisors, and security officials. Or take the situation in this chapter's introduction where a computer plays the role of customs officer. These roles remind people of known relational asymmetries between social actors and trigger accepting or compliant responses. Based on the psychological insights from the previous paragraphs, it can also lead them to interpret and respond to the interactions with a technology as an authority situation.

In our interactions with digital technologies there are all sorts of names and titles that serve as situational cues suggesting authority. A common example are technologies that are labelled "expert" or are given a name suggesting authority such as *Socrates* –a digital system for calculating social benefits (NOS, 2021) – or *COMPAS* – software used to predict future criminals (Angwin et al., 2016). Nass, Reeves and Leshner (1996) also experimented with the use of authoritative labels for digital technologies. In one experiment, a television - a technology that is not typically considered to have particular expertise – was labelled a "specialist" (Nass, Reeves & Leshner, 1996). Participants were randomly assigned to watch fragments of news and comedy shows on either "generalist" or "specialist" televisions. Those who watched the "generalist" TV were told it was an ordinary 20- inch screen used for both news and entertainment shows and had a sign that read "News and Entertainment Television". The other group - watching the "specialist" televisions - was told they would watch the shows on two different televisions: one television that was used exclusively for

news programmes and had a sign "News Television" on it, the other (identical) television set was situated next to the news TV and was said to be only used for watching entertainment shows. This TV carried the sign "Entertainment Television". After watching the various fragments, the participants were asked to evaluate the overall quality of what they had seen and whether they had liked it or not. It was found that the participants watching the "specialist" TVs thought that the news fragments were significantly better (more informative, interesting, and serious) and the entertainment fragments were significantly funnier and more relaxing than the participants watching the "generalist" television.

Digital technologies can become central in authority situations in similar ways as human authorities do. Minimal cues conveyed by the roles and labels of digital technologies can already leave an authoritative impression and easily trigger individuals to respond in accordance. Take also the experiment conducted by researchers from the Georgia Institute of Technology (Robinette et al., 2016). This experiment showed that in pressing situations humans are very willing to blindly rely on a robot in the role of emergency worker – even when the robot is not very reliable and looks more like a vacuum cleaner. The experiment started with the robot guiding a person to a meeting room. The robot was controlled by a hidden researcher who either lead the subjects to the meeting room directly or who first made his subjects enter an unrelated room and led them in circles before reaching the right destination. Sometimes, the robot even broke down entirely and one of the experimenters had to further instruct them where to go. Upon arrival, the participants were asked to fill in a survey and read an article on indoor navigation technologies. Then, without the participants aware this would happen, an emergency occurred. The experimenters filled the hallway with smoke and the alarms went off. The robot's red LEDs lit up the words "EMERGENCY GUIDE ROBOT" and everyone had to decide whether to follow it. As it turned out, all participants did. They did so even though half of the participants had already experienced the robot being unreliable in the navigation guidance task just minutes ago and despite the presence of a standard emergency sign pointing them towards a different exit (Robinette et al, 2016: 1). And even when the robot performed errors during the evacuation as well – for example by pointing the participants towards a room blocked by furniture with no discernable exit – the vast majority continued following the robot anyway.

Surely, digital technologies are also increasingly capable of using more sophisticated cues to convey images of authority. Think for example of technologies that manifest emotions or are highly interactive. Or consider technologies that display specific language styles or the visual representations of authority symbols - e.g., the show of certificates or logos. These are all elements that can play a role in the construction of authority situations with both human and digital authorities. However, the empirical evidence for the idea that more and better humanlike cues will trigger stronger social responses towards an inanimate object is limited (Reeves & Nass, 1996; Roubroeks, 2014). Ergo, digital technologies do not necessarily need to look or act in the exact same way as human authorities to be consequential.

5.7 A DIGITAL PROXY?

In this final paragraph let's look at a critique on the idea developed here. So far, I have argued that people can respond to digital technologies in very similar ways as to human authorities because, in essence, they tend to view them as social actors. When the ontological distinction between humans and digital technologies is obscured, it becomes less strange to see people accepting the output of a computer over their own judgement. Just like human experts and authority figures, information technologies can become authorities that possess the capacity to produce consequential communications, because people do not seem to differentiate that much in their interactions with people or digital when recognising certain marks of authority. Digital technologies seem to be capable of tapping into old sentiments and perceptions of appropriate responses to those who are deemed wiser or in particular institutional positions. The fact that their authority is not formally or institutionally ascribed to them does not seem to matter much. When the performance of a digital technology is recognised as authoritative, then people will respond to it in the same way as they do to human authorities: with attitudes of trust, respect or to pretend as if this were the case.

But what if our social responses to digital authorities are not actually directed at the technology itself, but rather at those "behind" it? – the idea that digital technologies are just a proxy for human authority. According to this idea, people perceive their interactions with computers essentially to be interactions with the machine's programmer - or some other person involved in the creation or design of the system. And since programmers, designers and their bosses are people, it is only logical for people to interact with computers in a social manner (Sundar and Nass, 2000; Nass and Moon, 2000). The technology itself then only functions as a medium embodying the responses of a machine's maker or owner.

While this idea does not challenge the idea of anthropomorphism - after all, the belief that one is interacting with a machine's maker is compatible with the conviction that the machine is human in a certain way -, it does challenge the CASA-paradigm. Sundar and Nass (2000) have tested this *technology by proxy* explanation. They performed an experiment in which two groups of participants had to work on tutoring tasks on different computers with diverging "teaching styles". For the first group of participants, the computer was exclusively referred to as a "programmer"; for the second group of participants, the computer was referred to as a "computer". Throughout the experiment dozens of references were made to emphasise that they were either interacting with a computer or a programmer. For example, during the instruction and practice session the participants were told that they "will be working with this [computer] [programmer] to learn about the United States" or "this [computer] [programmer] will provide you with 15 facts" (c.f. Moon & Nass, 2000: 95). Sundar and Nass hypothesised that if people treated the computer as proxy, there would be no difference between the two groups; "computer" and "programmer" would then be just synonyms. However, a difference was expected when people would treat computers as a distinct source to whom they directly respond. This was, indeed, the case. The researchers

found that the participants who were told they were interacting with a computer experienced the tutoring session as significantly more positive and effective, and thought the computer was more playful and more similar to their own style of teaching and evaluating than did the programmer participants (Sundar and Nass, 2000: 693). Sundar and Nass realised that a programmer constitutes only one of the possible persons a computer user could orient to. Therefore, in a later experiment, participants were told they were interacting with a human tutor sitting in another room ("a networker") instead of a programmer. Again, a difference was found between this group and the group that was told it was interacting with a computer.

This largely refutes the computer as proxy explanation. It is also in line with the responses given by the participants in virtually all other CASA experiments when asked whether they had the programmer or any other human in mind during the study. They unanimously denied this was the case. They also denied displaying social behaviors in the first place. This further refutes the proxy hypotheses – had they imagined interacting with the programmer, they would not have denied being social. Moreover, in those experiments in which multiple computers were used all participants believed that the same person had written all used programs. If the participants considered themselves to be interacting with one and the same programmer, one would expect no difference between the groups on different computers (Moon & Nass, 2000: 94). However, still significant differences were found between the different computers.

What anthropomorphism and technology by proxy share is that they both build on the premise that individuals' social behavior towards technology is *consistent* with their beliefs about the ontological status of technology (Nass et al., 1993: 544). In anthropomorphism, people's responses are claimed to be in line with their erroneous beliefs about the ascribed feelings and intentions of a technology. In the technology as proxy view, people's responses are seen as a logical consequence of their directedness towards the machine's human maker. Both explanations naturally agree that many normal, informed individuals do not apply social rules *directly* to technologies. The CASA research discussed here goes beyond this assumption. It argues that all individuals can be induced to treat computers as social actors, even when they do not have an imagined relationship with some person behind the machine or when they are fully aware of its non-humanness. Generally, the CASA paradigm asserts that people's social attributions are made *directly* at the machine, even though we know they are not human in any way (Fogg & Nass, 1997: 552). Nass and colleagues consider such responses to be *ethopoeic* (ibid., Nass, Steuer, Henriksen & Dryer, 1993; Nass, Steuer, Tauber & Reeder, 1993; Nass & Moon, 2000). *Ethopoeia*, as described by Moon and Nass (2000: 4), "involves a *direct* response to an entity as human while knowing that the entity does not warrant human treatment or attribution". In other words, *ethopoeia* refers to the process in which people treat computers in a social manner, despite its *inconsistency* with their espoused beliefs about machines (Nass et al.,1993: 544).

In conclusion, human and digital authorities share more similarities than we might realise. When interacting with digital technologies, we often forget that we are engaging with an object. We respond to digital technologies much in the same way as we do to human authorities when the technology's performance is perceived as sufficiently authoritative. This experience is shaped by the same symbols, positions, and roles associated with human authorities. The ability of digital technologies to issue communications that are recognised and responded to by human beings is a significant prerequisite for establishing digital authority. The fact that digital technologies cannot think for themselves, generally lack formal powers, and do not always warrant unquestionable acceptance does not stop us from relying on their output. Digital authority may not always be formally embedded, but it is real terms of its effects. And as was discussed in this paragraph, when people recognise digital authority, they are not directed at some other human authority or designer behind the technology. With digital authority an audience is directly focused on the technology itself.

5

6 CULTURE

In 2018, a Dutch newspaper article reported on the story of woman whose car was stolen in 1998 but remained registered in her name (NRC, 2018).[91] Because the license plates in the Dutch Vehicle Authority's (RDW) database were still linked to her, government agencies kept sending her letters – for example about the road tax that had to be paid and fines for the expired Periodic Technical Inspection (APK). The fact that she had reported to the police that her car was stolen had little effect. Officials kept pointing at the information system: the vehicle was registered under her name and thus she needed to pay. As the letters continued to arrive her debts with the government further increased. Debt collectors and police officers started visiting her house. She lost her job, and her social security benefits were withheld, leaving her unable to pay for her and her daughter's daily needs (Widlak & Peeters, 2018: 28). At some point she was diagnosed with rheumatism, allegedly triggered by her stressful situation. In 2011, after thirteen years, the vehicle was finally deregistered. However, the RDW refused to terminate the vehicle registration retroactively because this would risk the "purity of the system". Her debts did not disappear (ibid).

A man falls victim to identity fraud.[92] For years, his name was consistently used by an old classmate when committing crimes and as a result he became registered as a serious criminal, illegally residing in the Netherlands. The man started receiving court notices and all sorts of fines. Together with the police he then managed to solve the identity fraud. At least, that seemed to be the case. In the years that followed, he was arrested multiple times and at the airport he was refused to board his plane leaving him unable to travel. His home was also raided by 35 officers of the FIOD who accused him of money laundering.[93] It left the man physically, mentally, and financially strained. While the identity fraud was already known for years, the government struggled to clear the man's name as it remained unclear in which systems the wrong data were located. Officials continued to approach him with caution and their actions were still guided by what appeared on their screens. At some point the man was advised to take on another name. Only much later did his name get cleared, after almost twenty years of being a victim of identity theft.

These cases are extreme but perhaps they are more common than we think. Are they, with all the other examples from this book, maybe also a reflection of a broader

contemporary culture? The aim of this chapter is to shed light on what is easily kept in the background: the cultural layer underpinning authority. What makes digital authority claims *resonate* with an audience's values and principles? In chapter 3, I emphasised that authority is always grounded in audiences' culturally conditioned expectations. The authority relations and situations as presented in the previous chapters are shaped by cultural contexts. Below I discuss how contemporary cultures form ideas of what authority is, and in such a way facilitate the evolution of digital authority.

This chapter is divided into two parts. The first considers the differences between contemporary cultures. Paragraph 6.1 begins with the observation that people in different societies grow up with different ideas about authority. Paragraph 6.2, in turn, discusses the experience of technology across countries. These first paragraphs give rise to the question to what extent cultural ideas about authority and technology influence the potential for digital authority. After all, digital authority is not limited to those contexts where there is unalloyed enthusiasm about technological innovation or social hierarchies: even in environments where such attitudes are ambiguous there are plenty examples of people listening to digital technologies. In paragraph 6.3 attention is paid to this ambiguity by turning to the "tech fixes" – "tech ruins" binary: two contrasting narratives found in many contemporary cultures that leave space for a variety of conflicting technological behaviors and attitudes. This discussion nuances the idea that digital authority is confined to particular societies or cultural environments. In fact, one could even claim that individuals' views on technology are cultivated by both two extremes that, in their own way, carry a sense of awe and reverence in them.

The second part of this chapter focuses on what unites most forms of modern life. Paragraph 6.4 highlights our changed relationship to uncertainty and risk. Under conditions of modernity human beings have become increasingly aware of the contingent and hazardous nature of their world. It has led to profound changes in how we experience and act in the world around us. In paragraph 6.5, I argue that in modernity individuals and societies are engaged in questioning and reshaping their own certainties. As traditional authorities are challenged, there emerges a pressing need for new ones. Paragraph 6.6, in turn, concludes that the emergence of digital authority is best understood as a response to modern complexity and uncertainty. The final paragraph summarizes the main conclusions of part 2.

6.1 MENTAL SOFTWARE

> Stockholm December 23, 1988. The Swedish king Carl Gustav this week experienced considerable delay while shopping for Christmas presents for his children. When he wanted to pay for the presents by check he could not show his debit card. But the salesperson refused to accept the check without

> legitimation. Only when helpful bystanders dug in their pockets for one-crown pieces showing the face of the king did the salesperson decide to accept this for legitimation - but not, however, without testing the check thoroughly for authenticity and noting name and address of the holder.[94]

Authority is a fact of life and part of every culture. All around the world people need authorities to navigate their way through complexity. Yet, views on authority vary widely across the world. The renowned Dutch cultural psychologist Geert Hofstede has done extensive research on *culture* and "the patterns of thinking, feeling and potential acting that every person carries within him- or herself" (Hofstede et al., 2010: 4). To use a computer metaphor, Hofstede says culture works as our mental program or "software of the mind" (ibid: 5). This is not to say people are programmed like computers - a person can deviate from cultural patterns and react in ways that are novel or unexpected. In fact, cultures always come in *spectra* (or ranges) of acceptable or appropriate behaviors and there can be significant differences between individuals that allow for a variety of responses (Meyer, 2024: 29-31). Rather it is to highlight that, given one's past, certain responses become more likely and understandable. Much of our mental programs are learnt and come from the social environment in which we grew up and encounter a broad range of life experiences. They are acquired within the family – from parents who also acquired them when they were children -, at school, in the neighborhood, at the workplace and in the living community (ibid.). As Hofstede writes, culture is a collective phenomenon, covering the unwritten rules of the social game that are learnt in and shared with people who live *or* lived within the same social environment (Hofstede et al., 2010: 6). Culture, as he defines it, then becomes the "collective programming of the mind that distinguishes the members of one group or category of people from others" (ibid.).

Hofstede and his colleagues have spent much time researching cultural differences between countries. For his work at the human resources department of IBM, Hofstede traveled the globe in the 1960s and 1970s to interview people and conduct surveys on how people in large organisations behave and collaborate. The analysis of the more than 100,000 questionnaires resulted in "Hofstede's Dimensions": a widely used framework that highlights the effect of a society's culture on the values and behaviors of its members. One of these dimensions, for example, is *uncertainty avoidance*, which is about how well members of a culture can tolerate ambiguous or unknown situations (ibid.: 191). Do people embrace or avert unexpected events? It was found that a country such as Jamaica is more accustomed to ambiguity and tends to impose fewer regulations. In many Latin American and Latin European countries, on the other hand, such uncertainty causes more nervous stress and a need for written and unwritten rules to deal with unknown situations (ibid.: 192-4).[95]

Hofstede also investigated cultural views on authority. More specifically, he studied experiences of inequality and hierarchy in social relationships. Inequality can be found in

every society. There are always some people that are recognised as more wealthy, powerful or knowledgeable than others. Some are given more respect or status than others.[96] What is interesting is how people handle this fact. Hofstede focused on how, in different countries, inequality in social relationships is experienced, for instance between parents and children, teachers and students, bosses and employees, doctors and patients and officials and citizens. He asked people, for example, if they were comfortable disagreeing with their managers and whether they preferred their bosses to have an autocratic or a more deliberative decision-making style (ibid: 42). He found remarkable differences in the extent hierarchy in social relations is expected, accepted, or even preferred (ibid: 53-88). For instance, in countries such as France, Mexico, and China there is a significant gap or *distance* between people at different hierarchical levels and individuals often readily accept their position in a social hierarchy.[97] This is also reflected in organisational design with more hierarchical levels, managerial privileges, formal titles, and higher salaries.[98] In the United Kingdom, the Netherlands, and the Nordics this is much less the case; people are less concerned with status and have a more egalitarian outlook. In her bestseller, *The Culture Map* (2024), Erin Meyer observes a very similar pattern in how cultural preferences in leadership styles and formal hierarchical structures (Meyer, 2024: 1148-156). From this perspective, the salesperson's response from the beginning of this paragraph is less strange than it seemed.

In short, cultures differ in the extent to which people are socialised into following a hierarchy and recognising inequality in social relationships. This makes that in some places, authority requires little justification. Some cultures are more authority-centered or oriented – dictating a deep and lasting respect for authority figures. In other cultures, formal authorities may need to work (or perform) differently to be recognised as such. For example, in some cultures leaders may choose to renounce official symbols and gestures for the effective exercise of authority. For years, the former Dutch prime minister, Mark Rutte, typically commuted to his office by bike. This type of behavior of the powerful is very unlikely in Belgium or France, as also Hofstede concluded elsewhere (1980: 93-4). Still, the fact that authority operates differently across cultures does not seem to tell us much about the possibility of things as authorities. At most, it tells something about the different ways technologies could come to yield authority. How about general cultural attitudes towards technologies?

6.2 TECHNOLOGICAL CULTURES?

People are also socialised in different technological environments. The United Nations' *Technology and Innovation Report* series assess the state of innovation across different countries. It presents a "technological readiness index" based on indicators such as ICT deployment, skills, research and development activity, industrial capacity, and access to finance.[99] Unsurprisingly, in 2023, the ranking for 166 countries was dominated by high-

income economies, such as the United States, Sweden, Switzerland, and the Netherlands. In these countries, digital technologies are almost everywhere, and innovation forms an important societal driver. Countries in Latin America, the Caribbean and Sub-Saharan Africa are generally the least prepared to use and adopt such technologies (United Nations, 2023: xvii-xviii).

These studies are about technological *practices* or the implementation of certain technological ideas in society. They tell us something about the general extent to which people may be familiarised with digital technologies, but not so much about people's general orientations. Other publications teach us more about *values* and attitudes towards innovative technologies. For example, in one survey over 17,000 people from 17 countries were asked about their views on AI systems (Gillespie, et al., 2023). Here too, stark differences were found between countries. People in emerging economies such as India, China and Brazil reported the highest levels of trust, with most people trusting AI systems (Gillespie, et al., 2023: 3). In contrast, only a minority of people in Finland (16%), Japan (23%), Estonia (26%) or the Netherlands (29%) claimed to really trust AI (ibid: 13-8). Some countries clearly hold more positive attitudes towards AI than others, with people in France being the most fearful and even outraged at times (ibid: 20). But this is only part of the picture. Across countries, younger generations and the university educated showed a distinctly more positive orientation towards AI (ibid: 21, 51). And in all instances reliance on AI systems was found to depend on the specific application or use case. In every country there was a tendency to trust the use of AI in healthcare (i.e., for diagnoses and treatment decisions) and for public and safety and security decisions the most, and the use of AI in human resources (i.e., for aiding hiring and promotional decisions) the least (ibid: 13-5).

Perhaps most interesting about this survey is that it reveals that people commonly experience ambivalent feelings towards AI, reporting optimism and excitement on the one hand, while simultaneously reporting worry or fear about these technologies (Ibid: 19). In India, for instance, people are most likely to feel positive emotions about AI applications, but they also are most fearful and have one of the highest levels of outrage compared to other countries. It shows that in many countries fear and worry about AI coincides with optimism and excitement (ibid: 20). For the study's overall population, it was found two-thirds of the respondents feel optimistic about the use of AI, while about half also feels significant worries (ibid: 3).

We might even carefully extrapolate these cultural observations to all sorts of other digital technologies. Elsewhere, a broad study on trust and reliance on technologies also found that people in economically developing countries have a more positive outlook to digital technologies and a greater willingness to accept them at an earlier stage (Edelman, 2022: 11). And here too it seems that optimism and positive attitudes within countries tend to differ per technology (ibid: 41; Juma, 2016: 5) and that concerns about societal digitisation – for example about issues such as privacy and cybersecurity – are widespread (ibid: 14).

But what does this all mean? Ultimately, how do cultural ideas about authority and technology relate to the construction of digital authority? Based on the previous sections, one might hypothesise, for example, that there is only a limited cultural basis for digital authority in countries that are most "technologically ready" as they tend also to be the most concerned. But we have already witnessed the authority of technologies not being confined to certain societal contexts. Individuals from all over the globe have displayed a strong willingness to do as a machine told. This brings us back to the idea that cultural patterns come in spectra, and apparently there exists a wide range of acceptable attitudes or preferences across countries when it comes to technological innovation. This suggests that digital authority may emerge pretty much everywhere. Surely there are relevant differences in cultural orientations towards technologies that can promote or impede accepting attitudes and responses among people - a topic that deserves more scientific elaboration. But this does not tell us much about what happens in modern societies and how individuals can become oriented towards an idea of digital authority in environments where technologies are experienced as both a positive and a negative. Let's therefore zoom in a bit on two, contrasting, contemporary cultural narratives on technology that tend to simultaneously exist in many societies.

6.3 TECH FIXES, TECH RUINS

In most places, there is no single cultural take on technology. Instead, individuals often face at least two general conflicting narratives to which they must relate. On the one hand there are the many expressions of amazement of all the great things technologies have to offer. It is easily recognised that scientific and technological progress solved many problems that have tormented humanity for centuries - abolishing diseases, exploiting new energy resources, and offering new forms of communication and rapid transportation (Johnson & Wetmore, 2021:1). Technological development became a synonym for human progress and convenience, as well as an important ordering principle of modern states to deal with concerns and shape the world for good (Scott, 2020: 3-4).

The proliferation of *digital* technologies has further substantiated the conviction that technology will fix everything. Evgeny Mozorov (2013) refers to this contemporary tendency to solve a broad range of issues with digital technologies as "solutionism". It recasts every difficulty, no matter how complex, as something that can be ironed out by the right code, and algorithms - effectively making our world friction and trouble-free (Morozov, 2013: 5-6). Similarly, Meredith Broussard (2018: 14), labels "the belief that tech is always the solution" as "techno-chauvinism" and James Bidle (2018: 4) speaks of "computational thinking" to describe the "belief that any given problem can be solved by the application of computation". We find this self-confident logic that with innovative technologies, human beings can tame even the most complex and threatening problems

in many policy contexts. The World Economic Forum (WEF), for instance, published an article with the headline 8 ways AI can help save the planet (2018).[100] It describes "game changer" AI applications to "address climate change, deliver food, and water security, build sustainable cities, and protect biodiversity and human wellbeing" (WEF, 2018). Similarly, in the field of security the implementation of new technologies is often claimed a necessity to prevent crime and other unwanted behaviors (Byrne & Marx, 2011). Smart CCTV cameras, big data applications, iris scans, DNA profiling and other biometric technologies have all been introduced as indispensable to effectively deal with contemporary security issues.

But this thinking is much more widespread and goes well beyond policy ideas and initiatives alone. Studies have found that in children's books, technology is often depicted as autonomous, superior, and as something that fulfills human needs and wants (Axell & Bostrom, 2019: 32-5; Axell, 2017). There are commercial slogans like "Let's make things better" (*Philips*), "Sponsors of tomorrow" (*Intel*) and "Empowering us all" (*Microsoft*) – emphasising how technologies advance our lives. We read news headlines on how algorithms can better identify cancer (The Guardian, 2023) or about smartphones that spot early signs of Alzheimer's (The Daily mail, 2022). At the top of the charts are books that ask what Google would do (Jarvis, 2009) and highlight that "the future is better than you think" thanks to innovative tech (Diamandis & Kotler, 2012). In many movies and popular television shows, criminal cases are solved thanks to digital technologies.[101] In other words, techno-optimism, as mentioned in chapter 1, has many public shapes and forms.

There is another common narrative on technology in contemporary cultures that is radically different though, drawing people's attention to the darker side of technological development. The advert of science and technology has not just brought historically unparalleled levels of comfort and social security. It has also created new risks and the possibility of nuclear catastrophes and war, environmental degradation and climate change or bioengineered pandemics. They are anxiety-provoking because they are difficult to control and serious in their effects (Lupton, 2013: 16; Beck, 1992b: 101-2; Giddens, 2002: 34). Swedish philosopher Nick Bostrom has even argued that because of technological breakthroughs, "humankind may be rapidly approaching a critical phase in its career" (Bostrom, 2002:1). He writes that as our technological powers expand, so does the scale of potential consequences (Bostrom, 2013: 16). We may even have entered the era of so-called *existential risks* that could imperil human civilisation. For example, there have been several public warnings from experts that in the decades ahead powerful Artificial Intelligence might be developed that cannot be controlled and could result in human extinction or drastically curtail its potential. A survey among AI researchers found that a majority believed that there was at least 10% chance of an AI doomsday-type scenario (AI Impacts, 2022). In a recent essay in Time Magazine (2023), one of the founders of the field, Eliezer Yudkowsky, warned we need to shut it all down or "we are all going to die". Elsewhere, a similar message was shared by hundreds of experts and public figures who signed a

statement claiming AI poses an existential threat to humanity that "should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety, 2023; Washington Post, 2023).[102]

More generally, this critical narrative emphasises that technologies do not necessarily improve the human condition. Classical philosophers of technology long since worried about the role of technology in contemporary culture, warning that technology would alienate human beings from themselves and reality (Verbeek, 2005: 4). As Verbeek puts it, they feared that "people would come to exist only as cogs in the machine of a technologized society, reduced to the function they have in the apparatus of mass production, while reality would have meaning only as a heap of raw materials available to the human will to power" (Verbeek, 2011: 3). Above all, technology was approached as a monolithic phenomenon that is hostile to the human world. While the focus of philosophers has gradually shifted towards how different kinds of technological artifacts shape our everyday lives, worries about technological interventions to human life are still commonplace. Think for example of public and scholarly debates about surveillance or automated decision-making where technological solutions are easily portrayed as an undermining threat to human freedom and autonomy. Moreover, as argued by Morozov (2013), many things that are labelled as problems in need of a technological fix – such as inefficiency, ambiguity, and opacity – are often, after further investigation, not as problematic as they seem and may even be virtues in disguise that facilitate a life worth living (Morozov, 2013: xiii-xv). He maintains that some things just cannot be fixed – so called wicked problems that are characterised by ambiguity, complexity, and resistance to solution (Rittel & Webber, 1973) - or that in other cases a technological solution would make things worse for people. Not better.

The narrative that modern technologies are a "particularly dangerous enemy" to human beings and the environment is as ubiquitous as that of technologies as a staunch friend (Postman, 1993: xii). We learn about technological threats and failures through journalists and politicians. In schools, children are increasingly educated on all the different risks of the digital environment they grow up in (Livingstone, 2009; Lareki et al., 2017). And there seem to be as many television shows and movies confronting the anxieties around technology – from *2001: A Space Odyssey* to *Black Mirror* to *The Creator* where humans must confront future technologies - as there are productions applauding it.

What this previous discussion amounts to is that people in modern cultures often find themselves caught between these two conflicting narratives - between optimism portraying technologies as a solution and pessimism with regard to technologies' overwhelming and unstoppable power. Accusations fly back and forth, with techno-optimists being labelled reckless and naïve and techno-pessimists said to be conservative and outdated. But what these narratives share is a general deterministic approach to technology as the driver of change in society. Technological determinism is defined by Winner as "the idea that technological innovation is the basic cause of changes in society and that human beings have

little choice other than to sit back and watch this ineluctable process unfold" (Winner, 1986: 9-10). This kind of determinism is a strong claim, framing a causal relationship between technological development and societal change that leaves no room for human agency and inventiveness, neither in a role to curb technological development, nor to choose embracing it when it offers solutions to challenges that cannot be resolved without it. Having said that, this form of determinism is implicit in many societal discussions about digital technologies where innovative technologies are said to either fix or ruin the world. By accentuating these transformative capacities, technology turns into something potentially awe-inspiring. It creates a general atmosphere that can foster deferent attitudes towards digital technologies and that is hospitable to the construction of digital authority relationships.

At the same time, both narratives are essentially ideological rather than pragmatic. They are considerably different in how they portray the world and what actions are needed – inciting human beings to either follow or rebel against innovative technologies. But it is unclear how they guide human beings in their everyday lives and situations. However, recognising this duality in societies does help us in understanding the ambiguity in technological practices and attitudes that were highlighted earlier (paragraph 6.2). Being in a society with both strong positive and negative technological ideas could lead to contradictory behaviors when interacting with different technologies. In other words, many find themselves in a cultural environment that can both stimulate and limit digital authority.

Given these discrepancies, we perhaps should broaden our scope to better uncover the cultural foundations for digital authority. What if we turn to human beings themselves and how they experience the modern world? What is the role and place of authority here, and would this support the idea of digital technologies as authorities? This brings us to issues of uncertainty and risk.

6.4 A CULTURE OF RISK

Human beings have always lived under conditions of uncertainty. You can never know what life throws at you and what others will do. For a long time, people have coped with the dangers of uncertainty through religion and tradition. As Lupton (2013:2) writes, the supernatural was taken as a given, covering notions of both vengeful and testing gods, and evil demons (Lupton, 2013:2). Frightening events such as natural disasters, famines, and plagues were lived as something external, as "'strokes of fate' raining down on mankind from outside" (Beck, 1992b: 98). A system of superstition guided everyday life to ward off such evils – including beliefs in omens, offerings to the gods, totems and amulets, pilgrimages to shrines and the avoidance of tabooed places and people (Lupton, 2013: 2). Most dangers or uncertainties, however, were accepted as things over which humans had little control. At most, people hoped to indirectly influence the chances of a good life by calling on and pleasing higher powers.[103]

Over the centuries, human understanding of the world fundamentally changed. People began to view risks and misfortune more in terms of things that could be directly influenced by their decisions. The introduction of the concept of risk captures something of this changing mindset. It assumes that the future is not predetermined by a divine plan or will inevitably happen but is uncertain and open to human interference – the individual, alone or with others, can try to avoid or minimise harm (Breakwell, 2014:2). The origin of the word risk is often associated with early maritime ventures (Beck, 2003: 101; Giddens, 2002: 22; Breakwell, 2014: 2; Lupton, 2013: 5-6). Merchants commonly spoke of risks to denote the likelihood of their cargo reaching the port safely. Taking a risk meant embarking on a hazardous journey that could result in great commercial success and wealth, but also carried the possibility of shipwreck, disease or death.[104] To deal with the possibility of a storm, a flood or other perils that could jeopardise their voyage, people began to roughly assess the likelihood of such events happening and to look for ways to reduce their impact (Lupton, 2013: 6; Bernstein, 1996: 95). Trading companies then pooled their resources to manage the risk of potential losses and an insurance system developed. They discovered a method that subjected the future to human intervention. Probability calculations and statistics turned the ever-present danger of shipwrecking from an individual and unpredictable fate into an event that is generally calculable and shared between trading companies.

The emergence of modernity in seventeenth- and eighteenth-century Europe is typically linked to the further development of the idea of risk. Giddens (1991) refers to modernity as "the institutions and modes of behavior established first of all in post-feudal Europe, but which in the twentieth century increasingly have become world-historical in their impact" (Giddens, 1991:14-5). Roughly, modernity can be seen as the equivalent to the industrialised world, covering the process of industrialism, capitalism, institutions of surveillance and destructive weaponry (ibid: 15). With the entrance into modern society and Enlightenment thinking, a growing conviction developed that for human progress and social order objective knowledge of the world was needed. Through scientific study, reason and logic the laws of the natural and social world could be explored, measured and, ultimately, predicted (Lupton, 2013: 6). A calculating approach towards the world developed in the early modern era. Phenomena that were once viewed as striking individuals unexpectedly became quantifiable by considering them across populations. The consequences of events in a wide range of areas were now viewed as something broadly predictable and thus manageable by human decisions– as risks. While many threats were still considered as originating outside of human influence (Beck, 1992b: 99; Giddens, 1999:4; Giddens, 2002: 26) - although not necessarily viewed as being God's will but also as possibly coming from unmastered nature or unalterable life-situations – the logic developed that although, say, bad harvests, floods, sickness, unemployment could not always be prevented, at least their damage could be mitigated. By the 19th century, a comprehensive insurance system had developed, where

predictions based on probability calculations and the principle of "money for damages" made uncertainty about the future seem much more controllable (Beck, 1992b: 100).

Under the intensifying forces of modernity, a new phase emerged where risk became the defining characteristic of society. Beck (1992a: 19; 1998: 10) and Giddens (1991, 4; 1998: 26; 2003: 43) describe how "the end of nature" and the "end of tradition", together with developments in science and technology, gave rise to *risk society* or *risk culture*.

The end of nature does not refer to a world in which nature has disappeared, but rather to the idea that, in the last decades, human efforts of mastery have resulted in a situation where there is not much in the physical world untouched by human intervention (Giddens, 1991: 3). As Giddens points out, this gradual intrusion makes that people switch focus from all sorts of anxieties about what nature could do to them, to "what we have done to nature" (ibid.). Or as Beck (1998: 11) says, threats that were previously seen to originate from our external environment are now increasingly linked to human activities that inescapably contaminate nature. For example, floods are no longer regarded as purely natural phenomena but a consequence of human induced climate change (ibid.). And through technologies such as ultrasound and anticonception, having children is no longer something that is natural and external to us, but is invaded and subject to human decisions.

A society that lives after tradition, in turn, is essentially a world where people can no longer take traditional certainties for granted (Beck, 1998:10). Giddens explains that for a long time, many aspects of people's life were established by tradition as fate (Giddens, 1998: 26). During industrial society, life was pretty much about work and family which were typically organised around the idea that women were destined to be in a domestic environment for much of their life, have children and take care of the household. The fate of men was to work until they retired and to "essentially fade away" as Giddens puts it (ibid.). From the 1960s many of these given rules and securities about work, family and self-identity began to disappear, and people needed to make more decisions and choices themselves. With the end of tradition, Beck (1998:10) argues that we no longer live our life as fate, but now must find our own certainties in a process of individualisation and reflexivity (see next paragraph).

Both fundamental transformations in modern societies are related to the growing influence of science and technology (Giddens, 1998: 25). Advanced technologies are now central to our existence, and the things we feel and experience come increasingly under a scientific spotlight, as Giddens (1998: 23) puts it. At first this gave us the impression of more certainty about the world and a sense of control. And to a certain extent this was true – scientific knowledge, for example, led to cures for threatening diseases and higher life standards. But the situation was bound to turn against itself as science is an inherently skeptical endeavor, says Giddens, driven by a process of constant challenge and revision of claims to knowledge (Giddens, 1998: 23-4). Today's scientific claims are likely to be altered and developed in the future, contributing to a growing experienced complexity of

our environment. It makes it increasingly difficult for people to make predications in all sorts of situations. For many scientific innovations and experiments we cannot foresee the effects. Beck (1998:14) writes that society becomes a laboratory, where large-scale experiments are being conducted without anyone being responsible for the outcomes. To study the functioning and risks of nuclear technologies or find out about the theories and assumptions of test-tube babies, they have to be built or born – the established logic of scientific discovery that presupposes testing before putting into practice, where mistakes can be corrected, is lost in the age of risky technologies. Humans are now confronted with a range of new uncertainties and threats they have created themselves but cannot entirely grasp. Giddens therefore concludes that we now live "at the barbaric outer edge of modern technology" and on a "high technological frontier which absolutely no one completely understands, and which generates a diversity of possible futures" (Giddens, 1998: 25).

With tradition and nature no longer seen as a given, and with the intensification of science and technology beginning to obscure the future, a risk society develops - a high or late modern world where individuals are preoccupied with risk. This is not to say things are more dangerous than before. There are sufficient statistics available to chart out that many people are now in a much better position than in most previous ages (Rosling et al., 2019: 60-64; Pinker, 2018). Think for example of infant mortality rates, lifespan figures, homicide data and the proliferation of endemic illnesses which have all improved over the last centuries. Rather than to indicate an increase in the actual prevalence of life-threatening dangers, the term risk society refers to a change in how human beings perceive the world. People are increasingly aware of the complex and unfamiliar nature of modern life (Giddens, 1999:3-4). Beck (2003: 96) has described risk society as a "world out of control [where] there is nothing certain but uncertainty". It refers to a reality that appears ever more open and evolving – a world where all this new knowledge about our world has not left us with the increased sense of control Enlightenment philosophers envisioned. Risk society, in other words, involves living with contingency. It has made people increasingly oriented towards the future. To anticipate and control what will happen next, the future is consistently pulled into the present, and tinged with concern. In earlier stages of modernity, the future still also carried a positive connotation of opportunity. However, today's scrutiny tends to be more pessimistic, focused on identifying signs of danger (Jensen, 1997: 21).

6.5 UNCERTAINTY AND NEW AUTHORITIES

Modern life can feel inherently unsettling in a world characterised by contingency. Being liberated from the constraints of tradition and confronted with a flood of new information leads individuals to increasingly focus on creating and revising their certainties (Giddens, 1990: 38-9). The term *reflexivity* is often used to denote this response to complexity and uncertainty. Reflexivity is the process of constant critical assessment of social activity

and its contexts (Giddens, 1990: 36-7). While reflexive monitoring has always been part of human activity (Lupton, 2013: 99), modern reflexivity involves awareness of the contingent nature of knowledge and social action - that everything is temporary, incomplete, and contested (Giddens, 1991: 20). Given the uncertain future and permeance of change, the best preparation for human beings appears to be maintaining a level of flexibility – or *liquidity* as sociologist Zygmunt Bauman has put it elsewhere (Bauman, 2012:ix).[105] When making choices and for ontologically security, individuals now live in a reflexive reality where they will be continually monitoring their world and weighing up who to trust or listen to (Giddens, 1990: 38-9; Volkova & Pruel, 2019: 54).

In the past, the unpredictability of everyday life and the existence of phenomena felt to be beyond human control strengthened traditional authorities. Religion, the family system, and the local community were most significant for stabilising authority relations and establishing trust, respect, and strong normative standards in traditional contexts. While these cultures also harbored skeptics and magicians for those diverging from religious orthodoxy, these dissenting influences were hardly viable alternatives to the pervasive influence of tradition and the dominant religious system, which provided a guiding framework in many aspects of social life (Giddens, 1991: 194:5). They offered an alternative to doubt. In fact, individual experiences of living in a world full of external dangers were often actively nurtured by traditional authorities while presenting themselves as the only ones capable of understanding and successfully controlling these. Giddens for example speaks of religious authority creating "mysteries while simultaneously claiming to have privileged access to them" (Giddens, 1991: 195). This observation takes us back to the discussion in chapter 3 on authorities claiming to have differentiated access to something from which others are barred.

In modernity, in many areas of social life there are no longer such fixed or determinant authorities. While religion and tradition still have a place in today's world - and even may gain prominence in certain environments precisely because of modernity's complexity and doubt - there is a fundamental difference compared to earlier times: they have become authorities among others (Giddens, 1991: 194). In a reflexive society, guidance by tradition is largely replaced by expertise. In a world that is more open and offers more choices, people must constantly seek reliable knowledge of a variety of others "as guides to the risks that they run, the hidden dangers they face and the safest course of action available" (Garland, 2003: 74). In contrast to pre-modern and early modern times, people now cannot simply rely on local knowledges, tradition, and religious principles to conduct their everyday life. Rather, as Lupton writes, "they must look principally to experts they do not personally know and are unlikely ever to meet to supply them with guidelines" (Lupton, 2013: 100). Yet as we have seen, in conditions of modernity there is no such thing as unquestionable knowledge and expertise – things are never definitive and always potentially imperfect (Beck, 1992a: 155- 182; Lupton & Tulloch, 1998: 19). The more people learn about science, the more they find out about the things we do not know, and that science is fallible. In a context of imperfect knowledge, there

are increasingly competing knowledge claims (Beck, 1992a: 29, Giddens, 1991: 31-2; Jarvis, 2007: 32).[106] The fact that experts are bound to disagree becomes familiar to almost everyone (Giddens, 1994: 86). Imprecise predictions and contested claims make people cynical about the idea of inexorable human progress through rational knowledge and science, stimulating the continual reflexivity of individuals and institutions (Lupton, 2013: 99-100). Paradoxically, modern risk society is thus characterised by both an increasing individual need for experts and authorities to cope with uncertainty, while at the same time, a growing awareness of their potential fallibility and susceptibility to chronic revision and change (Giddens, 1990: 20).

In this situation, where risk is experienced as downloaded on individuals who now must reflexively decide whom to listen to when making decisions, a space has opened up for a plurality of new authorities to emerge. Barbara Adam (1996) already argued that while much has been said about the shift from traditional "external authorities" to the reflexive "internal authority" of the individual, there is another, equally identifiable tendency, towards a constant search for new and different "external" authorities that prevent individuals from being in doubt all the time (Adam, 1996: 139). Our mental software thus gets updated. We now, for example, see television stars, journalists and social media influencers playing authoritative roles in society. Furthermore, I contend below that digital technologies also fit into this modern pluralistic landscape of authority. For some people it is psychologically difficult to accept the existence of such a diversity of, often mutually conflicting, authorities, and they end up finding comfort in more overreaching frameworks of authority, submitting to one single authority whose rules and provisions provide guidance in most aspects in life (Giddens, 1991: 196). Others might become overwhelmed by choice or paralysed by doubt and withdraw from social life all together. However, most individuals navigate their way through modern-day life by, among other things, relying on a series of different and new authorities.

6.6 DIGITAL AUTHORITY AS CULTURAL PHENOMENON

Digital technologies have now emerged to address pressing questions that were once the exclusive domain of human authorities. For example, when it comes to the diagnosis or treatment of disease, people now access not only to religious figures or doctors but also comprehensive digital medical systems. To forecast weather conditions, one no longer need to rely on local wisdom, elders, or television weather presenters; instead, one can simply open a digital app. When technologies prove effective, or when established authorities fall short or give rise to doubts, they become an attractive alternative source to trust and respect.

Furthermore, digital technologies have also entered environments where authority was previously non-existent. Consider the case of parking enforcement, where officers who used to rely on human senses and memory now depend on digital output for their operations. And also in the context of future crime and the identification of potentially dangerous

individuals, predictive algorithms and risk assessment tools now provide directives to an extent that was previously unprecedented.

Digital technologies stand out because they are readily available to answer our questions, which has led to the possibility of a more intense form of authority. Instead of occasional guidance, these digital technologies enable instant and continuous guidance. When we are lost, instead of asking a local for directions, we receive direct instructions from our GPS. When children fail to understand something at school, they can consult an AI-teacher while doing their homework late in the evening. Where people used to have little choice but to rely solely on themselves or patiently wait for authoritative directives, this is no longer the case. In modern times, digital authority is more than just a potential new external authority among others; it also serves as an accessible and constant source of ontological security, offering human beings not only clarity *about* their world but also comfort *within* it.

Digital authority fills a void generated by a world that is inherently uncertain, where known authorities are constantly questioned and easily challenged. It is the result of an unceasing search for new forms of authority that prevent people from feeling overwhelmed or threated by their world. In modern risk culture, digital technologies are perceived to make the world a bit less overwhelming and a bit more predictable. They become important mediators between individuals and an overwhelming unknown future they seek to tame (Ericson & Haggerty, 1997: 6 c.f. Gray, 2003: 319). In their role as authorities, digital technologies can reduce everyday complexity for us and offer an enhanced sense of control. In other words, a culture of risk provides fertile ground for digital authority to emerge.[107]

The observation that digital authority is rooted in modern uncertainty has an interesting parallel with Hofstede's cultural dimensions discussed in the beginning of this chapter. Hofstede stated that uncertainty avoidance is the most influential dimension in determining technology adoption (Cardon & Marshall, 2008; Hofstede, 2001: 632-3). In environments where people are uncomfortable with unknown situations, it becomes more likely for them to embrace technology. Building on the previous sections, I will extend Hofstede's observation beyond its national scope and assert that what most societies in modernity share is a general growing discomfort with uncertainty and risk. While differences between countries may still exist in the extent to which this discomfort is evident, uncertainty and doubt are indicative of our time. Against this backdrop, as digital technologies become increasingly capable and present in various potential authority situations, trust and reliance on these technologies are likely to become an integral part of daily life for many.

In short, under modernity, there has been a shift of orientation towards risk, which can be detected at all levels: at a personal level, as a shift towards a more self-reflective attitude and an openness towards new (digital) authorities; and at a more societal level, as a shift in the organisation of public life and services. Routinisation and risk assessment in organisations are often considered a response to the contingent and uncertain nature of modern society. Digital technologies provide relevant guidance for organisations to deal

with all sorts of risks. They help shape organisational cultures in which professionals are socialised in working with and listening to digital technologies. This certainly involves organisations in the field of security, where potential consequences are experienced as pressing. When risks are perceived as significant, it becomes likely new authorities will be explored. Digital authority, then, is an indispensable response to the unsettling risks in contemporary societies.

6.7 CONCLUSION PART 2

The second part of this book further explored the idea of digital authority following authority's layered structure. We now find necessary elements for the potential construction of authority in human-technology relations, situations, and contemporary cultures. I described that digital technologies have gained the ability to issue communications to be followed or believed and can become the focus of our experience. It has given rise to *relations* with digital technologies that are potentially asymmetrical by nature. As with human authority, this relational asymmetry can be traced back to ideas of differentiated access to knowledge and power. I described that in our interactions with digital technologies, we can experience that the technology possesses superior access to either a wealth of data and algorithms that persuades us in advance or to some ascribed or programmed coercive force that latently secures compliance.

Similarly, digital authority *situations* can now emerge. The right authority cues present in the situation can lead individuals to engage in scripted behavior that stimulates compliant or accepting responses. As research shows, digital technologies are quite capable in tapping into those pre-existing social scripts and motivate individuals to do or think as they are told. The moment when in a situation digital technologies become associated with typical authority cues (symbols, titles, positions, etc.), could trigger similar responses as if it were with a real person. Distinctions between human and digital authorities then seem to become less relevant.

The cultural layer reveals how people are conditioned within societies towards the idea of digital authority. In many places, we find contrasting narratives of technological innovation that facilitate ambiguous attitudes and behaviors. Here, we may come to recognise digital authority in certain technologies and places. Zooming out reveals that digital authority is also inherently a product of modern cultures. Modernity has sparked a growing awareness of uncertainty of the world. As traditional authorities lose their legitimacy, we are now increasingly oriented towards new (digital) authorities to reduce everyday complexity.

The findings of part two are summarized in table 2 below. The last part of this book will now focus on the limits of digital authority: the moments where people do not, or no longer, recognise digital authority. It concludes with a discussion on the mitigation of the risks of digital authority.

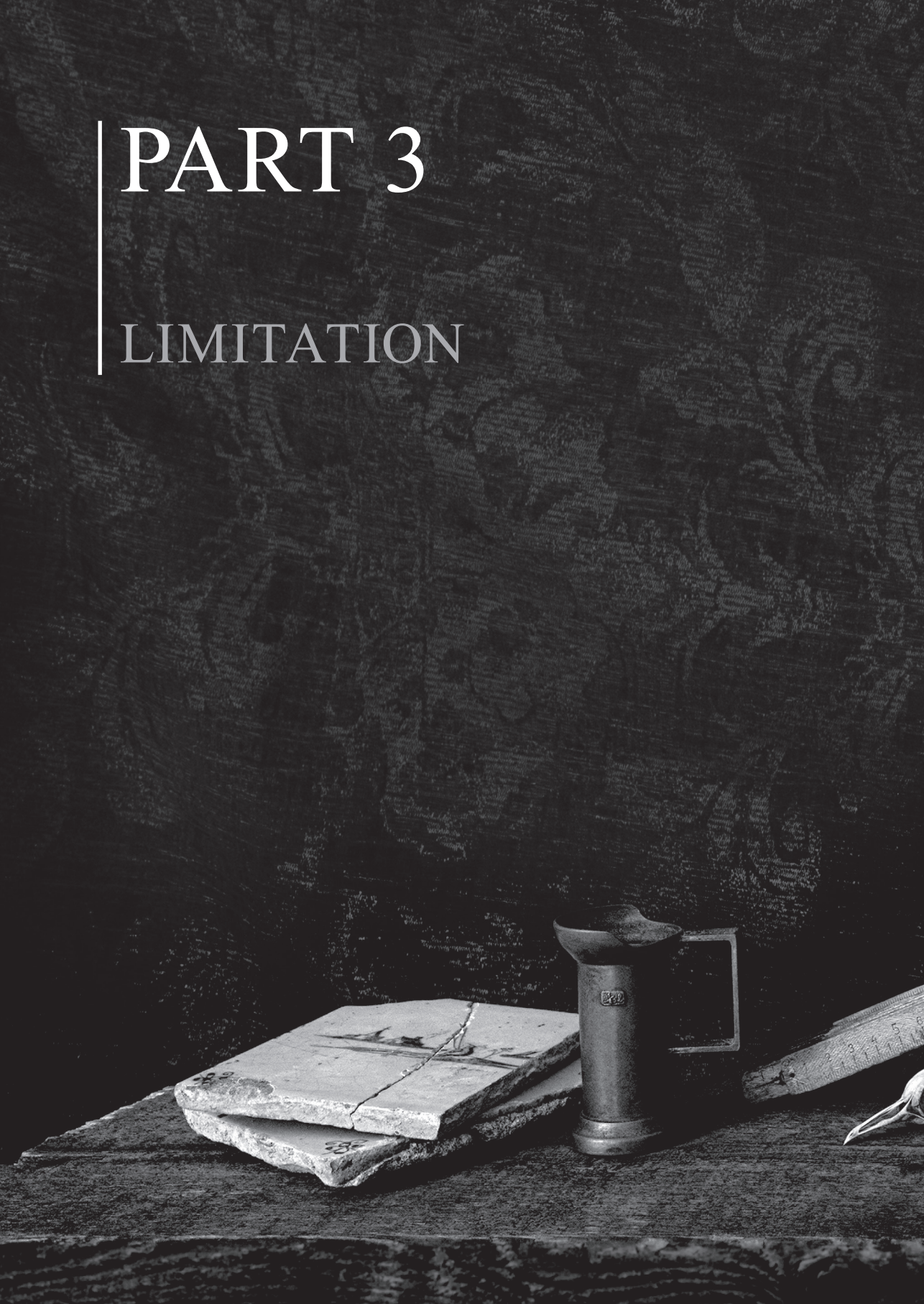
Table 2. Digital authority

	Relation	**Situation**	**Culture**
Characteristics	Experienced relational asymmetry	Defining an authority situation	Expectations of authority
	Consequential communications	Impressions	Conditioned values and beliefs
	Differentiated access to knowledge and force (persuasion and coercion)	Symbols (uniforms, titles, insignia, behaviors, places, etc.)	Dynamic and open to change
Digital Authority	Digital authority as asymmetrical human-technology relation	Digital technologies tapping into pre-existing authority scripts.	Narratives on authority and technology
	Digital communications to be believed or followed	Computers are social actors	Risk society and reflexivity
	Differentiated access to knowledge (epistemic niche/opacity data and algorithms)	Situational authority cues (positions, titles, symbols, behaviors, etc.) from digital technologies	Contingency and uncertainty facilitating openness to digital authority
	Digital access to force (authorisation/ technoregulation)		

6

PART 3

LIMITATION



7 HUMAN SAYS NO

Stanislav Petrov worked in *Serphukov-15* – a secret bunker southwest of Moscow and the nerve center of the Soviet Union's early-warning system for incoming ballistic missiles. His job was to monitor the system - code named Oko ("Eye") - and to directly report any alerts to his superiors. On 26 September 1983, just after midnight, alarms sounded. A red dot appeared on the electronic map covering the entire wall of the control room. It indicated that a missile was launched from an American base and was coming their way. "The siren howled very loudly, but I just sat there for a few seconds, staring at the screen with the word "LAUNCH" displayed in bold red letters," he told the BBC in 2013.[108] Not much later, the system picked up a second, a third, a fourth and a fifth nuclear missile. The red letters changed to "MISSILE STRIKE". The computers double-checked the data and displayed the reinforcing additive: "HIGH RELIABILITY". If this was true, the missiles would detonate on Soviet territory within 20 minutes.

Petrov had a choice. He could do what the Soviet military protocol dictated and report the incoming missiles to his superior. It was most likely a retaliatory strike would then follow, causing millions of deaths.[109] He could also declare the incoming missiles a false alarm and prevent a large-scale nuclear war, but with the risk that the Soviet Union would be critically hit if the missiles were real. Years later, Petrov described the choice he was facing:

> "My cozy armchair felt like a red-hot frying pan and my legs went limp. I felt like I couldn't even stand up. That's how nervous I was when I was taking this decision" (Barton, 2011).

Far from being certain, he decided to refute the computer's output and told his superior that there had been a system malfunction. A 50-50 guess he admitted (BBC, 2013b). "Got it", the officer - who, according to some accounts, was drunk at the time[110] – replied, and the message went up the chain (Hoffman, 2009: 11).

We know now that Petrov made the right call. After twenty-three minutes he realised that nothing had happened: "if there had been a real strike, then I would already have known, it was such a relief" (BBC, 2013b). It was found that the satellite had mistakenly

taken sunlight's reflection on a rare high-altitude cloud formation above North Dakota for a missile launch. "The sky always holds surprises for us," Petrov concluded (Spiegel, 2010). The algorithm for filtering out such data clearly needed revision and Petrov became known as the "man who saved the world".[111]

Doing what a digital technology tells us is important for the successful functioning of many things in our world – including safety and security. But it comes with serious problems when taken too far. As we have seen throughout this book, under the sway of digital authority people can be drawn well beyond the lines of reason and persistently follow technology into disaster. Fortunately, we do not always listen to these technologies. In fact, telling a computer "no" does not have to be difficult and it seems that many of us are all too familiar with situations in which we refuse to do or believe what the thing says. Still, while doing otherwise can be simple in theory, it might not be easy in practice. Especially when the stakes get higher and information scarcer, it seems an incredibly hard thing to do for everyone – as the story of Stanislav Petrov also illustrates.

The moment people do not accept the output of a computer signals the absence of authority and is illustrative of the idea that like human beings, digital technologies are not authorities by default, but only are when recognised as such by others. This chapter, the third part of the book, is about the limits of digital authority. In paragraph 7.1, I describe what it means when people say no to a digital technology. The following paragraphs, in turn, focus again on the relational, situational, and cultural layers of authority. The premise here is that the absence of authority can be traced back to elements in relations, situations, or cultures that obscure the construction of digital authority. Paragraph 7.2 argues, for example, that it is the absence of asymmetry in the relation between a technology and a human being that forms an explanation for people's dismissive attitudes and responses towards digital authority claims. In paragraph 7.3, I consider the use of different situational definitions and scripts that lead people to not recognise digital authorities. The cultural grounds for refuting digital authority are discussed, in turn, in paragraph 7.4. In the following section, paragraph 7.5, I focus on the moments in which digital authorities are challenged. What can digital technologies do to secure compliance among their human subjects in times of resistance? Paragraph 7.6 concludes this chapter with a discussion on how human beings and societies can be equipped to deal with the risk of erroneous or false digital authorities and how to mitigate their potential impact. Ultimately, dealing with digital authorities remains mainly a question of when one should rely on their communications and when to say no. The final paragraph presents a summary of part 3.

7.1 "I WILL NOT DO WHAT YOU TELL ME!"

The counter-intuitive and sometimes spooky things we can do under the sway of authority are clearly fascinating, and to date still much research is focused on authorities' potentially destructive speech. Among other things, the paradigm of human obedience to authority was transferred to the domain of human- computer interaction to test how far a person would go in torturing a robot (Rosalia et al, 2005; Bartneck & Hu, 2008). Still, people often do not listen to those claiming authority and may, for whatever reason, choose to think or act otherwise. And this holds true for many situations, including Milgram's experiments - in which a substantial part of the participants, at some point, refused to do as they were told (Miller, 1990; Packer, 2008) – as well as for our interactions with digital technologies.

The moment a message is questioned and fails to gain compliance constitutes the absence of authority. A quick observation of many of our daily interactions reveals that examples hereof are plenty. Not every teacher effortlessly silences a class, not every police officer can make a suspect cooperate without force, just like not every digital technology will be simply listened to by its users. Of course, drivers sometimes do end up in ditches when blindly relying on their GPS, but people also frequently ignore the digital navigational voices – with mixed success. The same will go for those working in a security setting; not every security official will automatically follow the recommendation of just any digital technology all the time.[112] Apparently users are not left to the whims of technology.

When someone rejects digital authority by saying "I do not believe it" or "I will not do it" it can reflect two things. First, it can indicate that authority was never there in the first place. For example, people might structurally ignore the output of a technology from the moment it gets implemented. Second, an audiences' refusal can also signal the negation of digital authority at a later stage. Here it is perhaps better to speak of the *deconstruction* of authority. People can, for example, at some point in time, come to question or dispute a digital technology's claim. Authority comes with great expectations, and when a technology fails to live up to these it will quickly disappear. While it is conceptually interesting to further concentrate on the failed construction and deconstruction of authority, for the purpose of this book I prefer a more general focus on the thing both have in common: the struggle of a digital technology to be consequential.

This rejection of digital authority, in turn, can manifest itself in various ways. It is most visible when people stop using a technology altogether. They will not even turn it on. Or they can simply end up ignoring it – the output of the digital system is not considered in human decision-making processes at all. This can be done *actively* (a technology's output is deliberately bypassed or openly refuted) or *passively* (the technology's output is not registered by the decision-maker in the first place). Digital communications can also be used in a more selective way. That is, individuals rely on their own observations and reasoning and only use the technology as an instrument for justification towards others. All these responses reflect a general rejection of digital authority.

Furthermore, rejecting, or challenging digital authority can happen at different levels too. It can be an individual response in which users become irresponsive to what a technology tells them. But it can also happen more collectively when peers start to openly call the technology into doubt. Such responses can be signaled through visible behavior revealing human defiance or in open verbal justifications for why the technology's communications must not be followed. In these situations, others rejecting digital authority can inspire individuals to do the same.

Last, as argued in chapters 2 and 4, sometimes it is (almost) impossible to work around or challenge digital systems effectively. Someone might try to say no or to reject their influence, but eventually some technologies will work their "will" regardless of what human subjects want. The code does not allow for it, the individual is not granted access to change the outcome. Still, when people do try to resist a digital technology, it signals the absence of authority. As a reminder, when people are forced to do as a technology tells them it is not authority but coercion. The rejection of authority it is not so much about whether people can act differently than a technology tells them, but more about their willingness to unquestionably rely on what they are digitally told.

Having established what it means to reject digital authority claims, let's now turn to *what* makes people not recognise digital authority. What makes us say or think: "no" when we interact with a digital technology? To answer this question, I draw focus again on the three layers constitutive of digital authority: the relation, the situation, and the culture in which needs and expectations of authority are shaped. When these layers do not coincide, digital authority claims will not resonate among its subjects. Below I investigate each layer to further understand such deviations.

7.2 RELATION: INDIVIDUAL AND TECHNOLOGICAL CHARACTERISTICS

Digital authority implies an asymmetrical relation between human beings and a digital technology in which the person feels the technology has a right to be listened to. This asymmetry is fundamental for all authority relations as it gives rise to trust, respect, and the willingness to accept whatever is being communicated. One way to understand the rejection of authority claims, in turn, is to consider it a consequence of the absence of such a relational asymmetry. Let's therefore take a closer look at the actors involved. Are there certain characteristics in people and technologies that frustrate the construction of such a hierarchical social structure between them? The following sections focus on the things that secure experiences of equality or even human superiority, in the relationships between people and technologies.

7.2.1 A matter of personality?

Psychological research on human authority has often focused on the personality traits of its subjects. And, indeed, it seems to make sense that people who easily accept and follow the words of proclaimed authorities have different personalities than those who do not. However, things are not so straightforward. One of the biggest insights from the obedience experiments was that the personalities of the obedient and disobedient participants were shockingly similar (pun intended). Elms and Milgram (1966), for instance, studied a variety of personality characteristics associated with obedience and defiance towards human authorities. They found very few differences between both groups. An exception was that disobedient participants scored significantly higher on "social responsibility".[113] It is plausible to think that those who have a low sense of social responsibility are more prone to obey an authority figure and shock an emotive victim. Conversely, someone who does feel responsible for what happens in his presence can reasonably be expected, at some point, to rebel against the one making certain claims or demands. Elms and Milgram (1966: 284-6) also found that authoritarian attitudes could predict obedience and disobedience. The participants that were fully obedient in the experiments consistently appeared to have more prior respect for authoritarian values than their disobedient peers. In other words, the research found a link found between peoples' authoritarian tendencies or personalities – as Theodor Adorno and colleague's (1950) called it - and their willingness to obey. This too is not much of a surprise, as also Bocchiaro and Zimbardo (2008: 30) observed: in a social situation with a clear hierarchical structure in which a lab-coated authority figure commands a research subject, you can expect that those with more respect for authoritarian values are most likely to do as they are told.

Elsewhere, the influence of moral reasoning on obedience in the Milgram experiments was discussed (Blass, 1991: 403). It was argued that with the moral development of individuals the likelihood of disobedience increased. And others also investigated dispositional factors such as social intelligence (Burley & McGuiness, 1977) and beliefs about the relationship between actions and outcomes – the extent to which one thinks one has personal control over one's life or that this is controlled by external forces such as fate or divine influence (Blass, 1991: 405). However, like with most other variables that have been tested, the outcomes were contested and deemed unreliable (Bocchiaro & Zimbardo, 2008: 31).[114]

What these psychological analyses suggest is that personalities are generally not the best predictors for individuals' decisions to challenge or refute authority. While it is intuitively attractive to assume that disobedience or non-compliance stems from a very particular set of personality factors there is not a lot of objective evidence for most such reasoning.

But what does this mean for our discussion here on the denial of digital authority and the relations between human beings and technologies? After all, the scope of our investigation does not just concern actors in formal positions of authority that explicitly command or order within an experimental setting but also include those that advice or suggest and operate in many everyday situations that could lead people to respond in

accepting or rejecting ways. Yet, studies into the role of personalities and the willingness to challenge technologies offer similar general insights. For example, Biros, Fields and Grunch (2003) found that people who are not optimistic about digital technologies in general are also less willing to rely on the output of a specific digital technology – a combat drone in their study. Szalma and Taylor (2011) found a negative correlation between neuroticism and agreement with a digital system (Szalma and Taylor, 2011: 82). Others found that introverts are less likely to listen to a computer than extraverts (Merritt & Ilgen, 2008). And Mcbride and colleagues (2012) showed that more intuitive nurses were more likely to accept the diagnoses from a digital expert system than nurses with a sensing personality.

But what this strand of literature also emphasises, above all, is that *personalities* are not a good predictor for human reliance on digital technology's communications. Other individual characteristics seem more relevant here. As discussed in Hoff and Bashir's (2015) systematic literature review on the issue of human-automation trust, *age* is often seen as a significant factor. But it differs in what way. Some studies have shown that younger adults, by default, trust digital decision aids less, but others, however, found that the more critical attitudes actually come with age (Hoff & Bashir, 2015: 414). Interesting in the context of this book is the study of Pak and colleagues (2012) who revealed that a picture of a physician in the interface of a digital diabetes decision support aid significantly enhanced younger participants' trust in the system's advice, but not that of older participants. Last, Hoff and Bashir (ibid) mention the role of *gender* in trust and reliance on digital technologies. But here too, researchers found no gender-based structural differences.[115]

These studies provide some interesting suggestions on the individual characteristics that could explain the differences in people's tendency to accept or reject digital authority. At the same time, this is exactly what they are: suggestions. Lack of research, scope and significant findings do not lead us to conclude that there is a typical user that tends to refute or ignore digital output and claims to authority.

7.2.2 Ascribed authority: Judging the system

There are all kinds of grounds on which people can feel equal to or better than a technology, and thus most likely to reject claims of digital authority. One is the institutionally ascribed authority of those interacting with the technology. Let me illustrate this with an example from the security domain of a digital technology that failed to acquire authority.

The use of digital risk assessment technologies in criminal justice is rapidly expanding across the world. According to Sonja Starr (2015: 205) we have entered the risk assessment era. Digital risk assessment here is about the use of algorithms that calculate "risk scores" for criminal defendants that estimate the chance that a person will commit crime in the future based on statistical inputs such as age, gender, criminal record, employment status and education level (Stevenson, 2018: 304). There is substantial support for the idea that these technologies are better at predicting future arrest than people and

are therefore increasingly used for determining restrictions on liberty – for example in decisions about length of a sentence, probation supervision levels and parole conditions.

In 2011 a law was passed in Kentucky making it mandatory for judges facing pretrial custody decisions to consult an algorithmic risk assessment tool that estimates the chance that a criminal defendant would flee or re-offend before his trial. Deciding about whether to detain or release someone while awaiting trial is a critical step in the justice process and getting it right is paramount for both the defendant and the community-at-large. The use of the risk assessment system was expected to make judges better at assessing who is safe to release (Mamallan, 2011:3). The effective release of low- and moderate-risk defendants would dramatically reduce incarceration rates without affecting public safety.

While the system was supposed to make Kentucky's criminal justice system fairer and cheaper this is not what happened. Judges refused to listen to the technology. After the law came into force, there was only a small increase in the number of suspects sent home. Most of the time judges simply decided to overrule the system's recommendations. If they would have deferred to the digital system, then non-financial release would have been granted to 90% of the defendants. But, in practice, as Megan Stevenson (2018: 373) found, this only happened in less than a third of the cases. In fact, it did not take long for judges to return to their previous ways and after a couple of years the release rate was even lower than before the law took effect.[116] And the story does not end here. Stevenson also discovered that after judges were required to consider the digital risk assessment system's output, white defendants were allowed to await trial at home without case bail much more often than black defendants.[117] In a different study using Kentucky data, Alex Albright (2019) showed that this was not simply a consequence of different risk scores by race. Rather, judges were more likely to override the system's default recommendation and to impose harsher bond conditions for black defendants than for similar white defendants.

The case of Kentucky confirms the idea that there is no such thing as authority by default. For Kentucky, digital technology was clearly not the magic bullet to improve the criminal justice system as it lacked the ability to produce consequential communications. Lawmakers seemed to know this already for some time. Before the 2011 law, the use of the digital risk assessment system was optional and most judges dismissed it (Stevenson, 2018: 343). But maybe, they reckoned, the system just needed a bit of help in establishing its authority? By making consideration of the technology's output mandatory, the new law was supposed to support the digital claim to authority by guiding the judges' attention to whatever the technology had to say. The support turned out to be insufficient though. Not digital authority but human discretion and bias kept guiding human decision-making.

But why is this so? In chapter 4 we have seen that digital systems could easily be recognised as authorities in contexts in which alternative sources of information are lacking (epistemic niches) – something that also seems to apply to the decision-making of Kentucky judges. There is only so much they can know about the risk of flight or future crime and

thus consistently doing what the computer says would have made perfect sense. Moreover, reliance on the technology, would have allowed them to work efficiently and to avoid the discomfort of (felt) responsibility (see also paragraph 8.4). When an inmate would reoffend, they could always point at the wisdom of the technology. Hence, why did the Kentucky judges did recognise the digital risk assessment tool's authority?

Stevenson (2018: 369-70) suggests that one of the reasons why judges only selectively incorporated digital risk assessments into their bail practices is that some might not have believed that these systems provided new and useful information: they did not recognise the system's expertise and thus did not experience the fundamental inequality of their relationship vis à vis the technology. This idea is also supported by a survey in which judicial attitudes to the provision of risk assessment technologies were investigated (Chanenson & Hyatt, 2016: 10-11). The survey indicated a general belief among judges that their judgement was more accurate than that of a risk assessment tool. In fact, less than 10% of the responding judges believed that the digital systems were indeed better at predicting future crime than themselves (ibid). The following section discusses the role of one's own perceived expertise in more detail.

Another ground for why these judges did not recognise the technology's authority has to do with their own position and the nature of the judicial office in general. As put by one author: "no other official is held to the standards of impartiality required of a judge" (Bedner, 2002: 14). Impartiality is fundamental to the judicial office and if judges fail to maintain it, their legitimacy will be damaged. Therefore, they are typically trained and frequently reminded to think for themselves, to make autonomous decisions. It is also anchored in their authoritative positions and the discretionary room granted to them. This authority in the executive sense (paragraph 3.2), translates into the relative autonomy to make decisions, using individual judgement or assessment when determinate rules are absent. As a Supreme Court order explicitly stated: "nothing in these guidelines shall be construed to limit the court's discretion as to whether or not to grant pretrial release to a defendant" (Supreme Court of Kentucky, 2011: §11). The result was that judges, indeed, frequently relied on their own authority to ignore the system's advice. The focus on their own authoritative role thus can also be a factor that prevents individuals' willingness to recognise digital authority.

The Kentucky case illustrates that real-world implementation of digital technologies can be different from what one might expect. Even when used in a context of limited information and time, and when its use is made mandatory by law, information systems can thus fail to gain compliance. When unrecognised by human authorities, digital systems remain technologies. Not authorities.

7.2.3 Expertise and system knowledge: wiser than computers

Authority relations often hinge on a sense of differentiated access: the idea that someone has relevant knowledge or experience from which oneself is barred. So, what if people come to hold such knowledge or insights themselves? Contrary to the static or enduring personality traits I discussed above (e.g., age, personality, gender), *domain-* and *system knowledge* are dynamic characteristics that people develop over time. Both are reasons for rejecting digital authority claims as they challenge the potential hierarchical structure of the human-technology relationship.

Domain knowledge or expertise is one of the most important reasons to say no to a digital technology. Expertise here refers to knowledge a person acquires over time. It is the result of extended experience in one specific domain. With greater expertise also often comes greater self-esteem or confidence, which is typically associated with a stronger tendency to rely on one's own judgement and resistance to influence attempts of others (Sagarin & Wood, 2007: 329). Research has shown that people with greater domain expertise are less likely to depend on digital technologies than novice users do (Sanchez, et al, 2014).[118] And it is also not difficult to come up with daily life situations in which our expertise leads us to refute the output of a technology. Think for example of drivers ignoring their GPS navigation in the area they live in because they know a shortcut or a medical specialist not needing an expert system to diagnose a patient because she has "seen this disease already a thousand times". Ergo, when people (think they) know things themselves they need the technology less and expertise provides a strong ground for such feelings.

In addition to expertise, we must also consider individuals' knowledge about a particular digital system. Having *system knowledge* is relevant for determining the limits and weaknesses of a digital system. It makes users aware of what a system can and cannot do and of its potential weak spots. Such knowledge can stimulate people to remain wary of unthinking compliance. As an illustration we can use Stanislav Petrov's life-or-death choice. Petrov "had a funny feeling in [his] gut" he told the Washington Post in 1999. The Soviet defenses were extensively briefed about a possible U.S. all-out nuclear attack – a single blow that would leave them completely paralysed. So, when five red dots appeared on the electronic map, doubts arose. "When people start a war, they do not start it with only five missiles (..), you can do little damage with just five missiles," Petrov reasoned based on his military expertise he had built during his twenty-six years of military service (Washington Post, 1999). But just as important seemed his technical background and knowledge of the digital system. As described elsewhere, Petrov was more of an engineer than a soldier:

> "He liked the logic of writing formulas, often used English-based computer languages. On most days, he was not in the commander's chair but at a desk in a nearby building, working as an analyst, responding to glitches, fine-tuning the software. Only twice a month he took an operations shift in order to keep

> on top of the system. When Petrov arrived eleven years earlier, the station was new, with equipment still in crates and the room empty. Now it had grown into a bristling electronic nerve center" (Hoffman, 2009: 7).

As the deputy head of the combat algorithms department at Serpukhov-15 Petrov knew the system's algorithms like no other; he designed the software and operating manual himself. He also experienced how the system was rushed into service as a response to the United States' introduction of a similar system but that the Soviet system was still troubled: shortcomings will be fixed later they were told.[119] Still, the electronic boards showing a missile launch never happened before. One of the things that made him suspicious was the speed in which the data was processed through the 30 levels of verification after the target was identified; Petrov doubted if the system could do that. He also considered various potential system malfunctions that could explain the alert. As he later said in interview with the German magazine Der Spiegel (2010): "We are wiser than the computers. We have created them". His system knowledge surely contributed to him not following the computer's output in an automatic and unthinking fashion. Petrov has always remained humble every time he was called "the man who saved the world". "I just did my job", he told a journalist (BBC, 2013b), "but they were lucky it was me on shift that night and not a dulled soldier". The latter, according to Petrov, would have probably followed the protocol and the digital technology rather than reasoned for himself.

Knowing a technology or the domain in which it operates helps to transcend the mysterious air that is constructive of the asymmetry that is endemic to authority relations. It is important to maintain a critical attitude that allows users to question digital authority and say no. This also goes for past negative experiences one can have with a (similar) digital technology (Manzey et al., 2012). Past failures, errors, and other malfunctions can also be grouped under the system knowledge header; when we have experienced a technology failing, it provides users with relevant lessons about its weaknesses that might prevent blind or automatic deference. This brings us to the next section about the performance of a digital technology itself.

7.2.4 Technological failures and inconveniences

When a digital technology fails to live up to our expectations, we are likely to reject its output. This often relates to the objective performance of the digital technology. When a technology does not provide consistent output – the technology is unreliable – or does not do what it is claimed to do – the technology is invalid – it impacts how we relate to it. This also goes for frequent error messages and breakdowns (Merritt & Ilgen, 2008). When digital technologies fail to perform authoritatively in the eyes of their human audience – e.g., they are caught being wrong or unreliable - compliance becomes highly uncertain. It seems that for digital authorities the adage also holds true that first impressions matter. Failures that

are experienced in the early stages of interaction tend to have a greater negative impact on individuals' willingness to rely on a technology than errors that occur later (Manzey et al., 2012). And especially "easy failures" seem detrimental to authority. After all, why should someone blindly rely on a technology for a complex and important task when it cannot perform its basic functions well?[120] The argument goes that when digital technologies fail or break down, the structure of the relationship changes. People are confronted with the technology's fallibility and warned that automatic and unthinking reliance is unwarranted.

A digital system that offers no clear advantages but only seems to make a task more arduous will also not convince an audience to do as it is told (Parkes, 2012). Indeed, this also ties in with the previous paragraph on expertise and knowledge. When people realise that a technology "is nothing special" and following it is in no way beneficial to them or others but will only generate extra work, digital authority claims surely become problematic.

7.3 SITUATION: SIGNIFICANT OTHERS AND WRONG CUES

Digital authority lies vested in the character of the situation. When people enter a new situation, they face the question: what is going on and what is appropriate here? In chapter 5, I argued that the right situational cues can easily trigger people to define situations with digital technologies as authority situations in which accepting or submissive responses are warranted. When digital authority is rejected, in turn, it can be considered the opposite outcome of this situational analysis – the setting a person entered did not lead him or her to experience it as if he or she was dealing with a digital authority. By contrast, the individual determined something is not an authority situation in the first place or rather recognised the authority of another actor present. Below I discuss some of the potential mechanisms that link to such responses.

7.3.1 The wrong impression

One way to understand an actor's failure to be consequential is as a sign of the absence of relevant situational cues that suggest authority. When thinking of technologies, this could bring us to the issue of *design*. For example, it is suggested that when system interfaces look less attractive, people are less willing to rely on them (Hoff & Bashir, 2015: 422). Or think of situations where the wrong symbols are used. When the portrayed insignia, logos or other things intended to mark the authority of a technology are not recognised as such - for example because they have become controversial or do not match the expectations of an audience – they will not be able to yield authority effectively.

In chapter 5 the idea of anthropomorphism was also discussed. Humanness in digital technologies can provoke feelings such as trust and respect that are important for reliance on digital technologies (Pak et al., 2012; De Visser et al., 2012). It can even stimulate trust resilience in people – their trust does not decline as much when system

errors occur. However, when designers do not consider individuals' expectations of a technology's interface, the result can be quite different. As mentioned earlier, when done wrong, designing humanlike technologies produces uncanny feelings and will frustrate digital authority claims instead of supporting them (Mori, 1970; Nass & Gong, 1999). In fact, it can be argued that it is impossible to perfectly mimic human authorities, and that every attempt to do so is doomed to fail as it will essentially reveal the technologies "non-humanness" (Nass & Moon, 2000). This could remind people that they are interacting with an object, and that the unquestioning acceptance of technological output is silly. We realise that an authority script is not appropriate here and get snapped back to real-life as it were.

This also goes for the *experienced* performance of a technology. There are many studies that highlight the positive relationship between the usability of a technology and human compliance (Hoff & Bashir, 2015: 422-3). It was found that when humans know how to read or interact with a technology, they become more likely to rely on it. Wang, Jamieson and Hollands (2011), for instance, describe that users were more willing to follow a combat identification system when it presented its output more explicitly.[121] One way to interpret this is that such explicit communication forms an authority cue that incites automatic acceptance among human subjects. But when communications are vague or unclear, it might not predispose an audience to respond with such accepting or obedient attitudes. The situational cues emitted may not to be associated with authority and might therefore not trigger the relevant automatic scripts and situational definitions. When a technology gets to complex or its output to incomprehensible, it might only lead people to think for themselves.

More generally, research on human-computer interaction suggests that communication style is an important instrument for digital authorities. Digital technologies can be endowed with a wide variety of "personalities" and sometimes there is a mismatch between an artificial personality and its human user that will work against acquiring authority. Whereas "polite" (Nass, 2004; Spain & Madhavan, 2009), "patient" and "non-interruptive" (Parasuram & Miller, 2004) computers can be good in motivating users to accept their output (for example in medical diagnostics), other personality types can have the opposite effect (Nass & Lee, 2001). Therefore, with digital technologies, too, it is not just about *when* they perform (results), but also *how* they perform that makes whether they can shape an authority situation for us.

7.3.2 Significant others

The behavior of others present in the situation can strongly influence how we end up reading a situation and thus our potential recognition or rejection of digital authority. In his studies, Milgram also demonstrated the importance of the social environment for disobedient responses to authorities. For one of his experiments, he used two rebel peers (1974:116-121). Both were actors, instructed to disobey at some point. The moment the victim "receiving" the shocks began to protest, one of these actors announced that he no

longer wanted to participate. Some trials later, the second actor withdrew as well. All tasks were now assigned to the naïve participant, as the experimenter had repeatedly insisted that the experiment had to continue. It was found that only ten percent of the participants in this variation of the experiment continued to obey to the end while in the initial condition this was more than 60 percent. While more research is needed here, a similar pattern seems to exist when thinking about digital authority; when people observe or hear strong negative sentiments among their peers against a digital technology, they themselves are also likely experiencing them as well. This was for example found by Michael Workmann who focused on subjective norms that "reflect an individual's assessment of social influences from important others such as managers and coworkers, about performing a behavior" – e.g., "my colleagues think I should not do X" (Workmann, 2005:213). Workman found that someone's perception of significant others' tacit and explicit views about a digital system significantly shaped their willingness to listen to a digital technology themselves. That is, the words and actions of others shape how individuals define and respond to situations with digital technologies. The dismissive acts or thoughts of peers are cues that affect individuals' ideas of what is going on and what actions are appropriate. It could lead them to conclude that rather than unquestioning acceptance, it is deviance that is warranted here. When thinking about any form of authority, audiences always matter.

Moreover, the *salience of a subject* - the person or thing about whom a decision is taken - can also influence people's willingness to follow a digital technology. Those who are affected by output of a computer are often distant or out of sight completely. They remain names or numbers on a screen such as is the case with many databases or mere images with a warning sign. Digital technologies have the tendency to *neutralise* their human subjects. But those who are affected by a computer's output can also be right in front of us, or we can easily identify ourselves with them. Under these circumstances doing or thinking what the technology tells us can become quite difficult. This can also be derived from the Milgram experiments where it was found that individuals' willingness to follow an authority figure depends on how clearly the negative consequences are presented (1974: 34-55). In the experiment's baseline setting, the participants were sitting in an adjacent room. While the complaints of the victim could be heard through the wall, they could not see him. In one variation of the experiment, Milgram placed the victim only a few feet away, in the same room as his subjects of study. With victims that could now be heard and seen, the level of obedience dropped from 62,5 percent in the initial condition to 40 percent (ibid).[122] These findings suggest that it is easier to follow an authority when the victims are out of sight. It is easier to push a button to kill a distant target than to kill someone face to face. Distance allows people to ignore, or even forget, that those who might be negatively impacted by their actions are real human beings. Not only are individuals more likely to respond empathically when they see a fellow human being in distress, moving a subject closer also confronts them with the causality between their actions and the negative impact on another. People are more likely

to reject authority when they see themselves as personally causing the harm that follows from doing as they were told.[123] And while it is true that with digital technologies those who are affected often remain distant subjects, making the impact of a decision largely invisible, even in these situations victims can gain salience for example by overtly complaining or by making themselves visible to the human decision-maker. Digital neutralisation can be countered and subjects can incite human decision-makers to follow their own judgement instead that of a technology. It confronts the idea that one is in an authority situation and draws attention to the idea that it is in fact appropriate within that setting to rely on one's own judgement.

7.2.3 Whose authority?

Sometimes we enter a situation with multiple authority claimants. The question then becomes not so much whether we define something as an authority situation, but rather whose authority are we talking about? Especially when those involved end up making opposite claims this poses an important issue.

There are many situations in which human authorities disagree, such as when parents are making opposite demands to their children or experts disagree over some issue. People then face the choice who to listen to. Similarly, there can also be situations in which a human authority figure diametrically opposes a digital technology. Depending on the setting, people may very well decide or be actively motivated to listen to a human authority instead of a technological one.[124] And sometimes it is the presence of other digital technologies that forces people to choose between two potential digital authorities. Consider the following case from the field of security.

In 1988, with the end of the Iran-Iraq war nearing, the American Missile cruiser USS Vincennes shot down an Iranian passenger jet that passed the strait of Hormuz, between Iran and the United Arab Emirates. All 290 people on board of the Airbus A300 were killed, including 66 children. The cruiser was equipped with an *Aegis system* - one of the most sophisticated computerised defensive weapons systems used in combat at that time – which had wrongly identified the aircraft as an attack plane of the Iranian Air Force. While the radar systems of nearby warships had correctly identified the flight as civilian, they all decided to rely on Aegis' judgement which they recognised as superior. As one of the crewmembers of a nearby ship later explained: "We called her *Robocruiser*, she always seemed to have a picture" (c.f. Noorman 2023).

What happened with the USS Vincennes is that multiple systems provided diverging communications about what was the case. This output could by no means be checked by the military officials: human beings lack the senses to gather relevant information as well as the brain capacity to comprehend all the input they received under conditions of time pressure. They listened to the Aegis system as it was generally considered to be the most sophisticated and effective system to track and target incoming missiles and airplanes. Yet, there were some grounds for doubt. The Aegis system was designed to detect targets at

great range, but consequently was unable to distinguish the size of a target. This entailed that the USS Vincennes was best suited for patrolling an open sea environment. But while the cruiser was initially deployed to the Western Pacific and the Indian Ocean, the USS Vincennes was later dispatched to the Persian Gulf to patrol its confined waters during the Iran-Iraq war in the 1980s. In this environment, however, the lacking ability of the Aegis system to determine size became problematic and contributed to the disastrous outcome. As it was put in a *Newsweek* magazine article:

> "In combat on the open sea, size doesn't matter much; a small plane or missile can sink a ship. But in the Gulf, where big civilian planes share airspace with small warplanes, size can be a crucial clue." (c.f. Dotterway, 1992: 140).

With the ongoing digitisation of their world individuals will be increasingly confronted with competing authority claims made by different machines. Situations in which different technologies present competing authority claims raise the question what makes people recognise the authority of one technology over another. For example, is new always better? New technologies can also come with problems or negative associations – for example when an earlier technology is deemed unreliable it could also negatively impact expectations of other technologies that could perform the same task (Ross, 2008). There are still many interesting avenues for further research here. For now, let us conclude that within a particular setting, technologies sometimes end up competing for digital authority and that it is up to human beings to define the situation and decide which gets recognised as such.

7.4 CULTURE: PROGRAMMED RELUCTANCE AND UNCERTAINTY

Cultures shape people's ideas and expectations of authority. They condition expectations of authority and technologies and influence which digital authority claims resonate within a given audience or, by contrast, will be refuted. Several cultural grounds exist that could lead individuals to challenge a digital authority.

7.4.1 Hospitable cultures

In chapter 6 it was already discussed that while authority has a place in every culture, there are considerable differences in expectations towards authority and ideas about the effective exercise of authority. It was suggested that across cultures, those in authority are often performing in very different ways. In some cultures, where individuals are socialised into following hierarchies and recognising inequalities in social relations, a strong emphasis on official symbols and formal statuses might help to effectively yield authority. However, in other environments such strategies can be counterproductive, and a more egalitarian approach towards an audience could be more successful. Different cultures may come with

different ideas about what is authoritative and what is not. One way to consider the absence of digital authority, then, is as the consequence of a disjuncture between a technology's expressions and individuals' culturally conditioned expectations.

There is also research that specifically focuses on the influence of culture on human-computer interactions in less abstract terms, highlighting that in some places people are, in general, less inclined to rely on digital technologies in the first place. One study, for example, found that Americans were less willing to listen to a decision aid compared to Mexicans (Huerta, Glandan & Petrides, 2012, c.f. Hoff & Bashir, 2015: 413). Elsewhere, significant cultural differences were found in how people respond to social robots (Hoff & Bashir, 2015: 413). The idea that some cultures are more open to technological innovation and display a greater general willingness to listen to digital technologies was already discussed in the previous chapter. While it was claimed that digital authorities can be found everywhere and that most cultures are characterised by a certain ambiguity towards technological development, it may very well be true that some cultural contexts are more hospitable to the emergence of digital authority relationships than others. It would be interesting for future studies to further investigate this issue and learn more about cultural differences in saying no to digital technologies.

7.4.2 Cultural unease and scandals

People's hesitance to recognise digital authority can also be rooted in societal experiences with the downsides and risks of science and technology. Common examples are disasters such as in Chernobyl (1986) and Fukushima (2011) that have strongly shaped contemporary perceptions of nuclear technology. Similarly, environmental degradation and climate change have become strongly associated with modern polluting technologies. Technological innovation has given rise to security risks that can be both global in reach and global in scale (Western, 2016: 100). It contributes to a general climate of technological unease, making people wary of technological innovation and complicating the construction of digital authority across cultures.

Critical incidents and negative experiences can also have a more specific cultural impact. This was also suggested in an article on the risk of algorithmic bias (Alon-Barkat & Busuioc, 2022). Alon-Barkat and Busuioc investigated reliance on algorithmic advice. Their study was conducted in the direct aftermath of a major scandal in the Netherlands, involving public authorities' overreliance on an algorithm with discriminatory effects – the "childcare benefits scandal" (*toeslagenaffaire* in Dutch).[125] They argued that the scandal resulted in an enhanced awareness of discrimination and algorithmic biases and a growing reluctance among people in the Netherlands to trust algorithms in its aftermath. It suggests that scandals can have a learning effect in societies – longer or shorter term – and, as such, influence the potential of digital authority.

7.4.3 Reflexivity

In the previous chapter it was argued that we now live in a reflexive reality where people are constantly monitoring and revising their world. In this world where tradition is challenged and under the growing influence of science and technology, everything has become more open and complex. People will constantly seek reliable information of others. This transformation has generated new authorities. To navigate their way through daily life, individuals rely on all sorts of authorities, including digital technologies that make the world a bit less uncertain and a bit more predictable. But these new authorities, too, are susceptible to potential revision and change as individuals are constantly integrating new knowledge to decide who to trust or listen to. While guidance by tradition has largely been replaced by guidance by expertise, in modernity there is no such thing as unquestionable knowledge – as discussed in paragraph 6.5. It is the thing that is fundamental to scientific and technological progress and when people learn more about science, they learn more about the idea that claims to knowledge can always be challenged and changed. There will always be competing knowledge claims or imprecise predictions, stimulating further the continual reflexivity of individuals and institutions. Hence, technological authorities are also fragile, in the sense that they too can always be challenged and replaced.

7.4.4 Organisational norms

Organisations often come with a distinct set of ideas about what is important, how things work and norms that regulate practices and direct organisational members towards certain goals (Reason, 1997: 192; Cooper, 2000; 112). The culture of an organisation influences how people experience their work and interact with other people. This can also include conditioning people's ideas and expectations of (digital) authority. As argued in the previous chapter, the contingent and uncertain nature of modernity can lead organisations to obligate or socialise people in working with and listening to digital technologies. But the opposite can also be the case. Organisations can design all sorts of mechanisms to ensure responsible use of digital technologies and develop shared perceptions on organisational security. This for example includes implementing structures that facilitate independent thinking and reflection by professionals or by placing emphasis on human responsibility. As such, organisational cultures can both offer a ground for the construction of digital authority and for its deconstruction.

One recent study of Meijer, Lorenz and Wessels (2021) explored the use of predictive policing technologies in Berlin and Amsterdam. They found that in Berlin, the use of a digital system for the identification of crime hotspots was facilitated by "linking the acceptance of the predictive policing system as a valid source of information, to the promotion of the norm that usage of the information system is good police practice" (ibid: 841). Here, the higher organisational levels emphasised that the system needs to be used and portrayed it as a form of rationalising the organisation. Police officers stated that they

felt that resistance was futile, and that they had little alternative than to use the system (ibid). In contrast, in Amsterdam the technology was explicitly portrayed as a system that can certainly help police officers in their practices, but that these professionals must "seek interaction between their own insights and knowledge, and the information from [the technology]" (ibid: 843). In short, organisational norms for using a system can be quite different and can influence the extent to which people willingly rely on it.

7.5 ARTIFICIAL UNINTELLIGENCE

What can be done when digital authority is rejected? What happens when subjects reveal they do not want to listen and do not recognise a technologies' authority? There are several ways in which compliance can be ensured. A common initial response of human authorities is to resolutely repeat and reiterate what was said – potentially supplemented with a raised voice or a changed posture. This is what parents do ("I'm telling you to go to bed!") and what Stanley Milgram did when confronted with hesitant participants ("It is absolutely essential that we continue" and "You have no other choice" (Milgram, 1973: 63). And it is also the default response of most digital technologies when people do not comply. GPS-systems repeatedly tell drivers to turn around, just like a digital system might tell its users that someone is a risk every time a particular name is reentered. Reemphasising one's words or message can be a convincing strategy for reminding an audience of the unequal relationship between speaker and listener and to ensure compliance.

But sometimes more is needed (and sometimes it is really good that people do not listen). Some respond to resistance with statements that make them sound interesting or stimulating to listen to such as managers or politicians reminding an audience of their great achievements or future ideals. Others might start making compliments or aim to lead by example. When challenged authorities can also begin to explain themselves to convince an audience or remind people of the great benefits of compliance. Similarly, for people in positions of authority there is also often the option of coercing a person into doing something – a manager can fire, a police officer can arrest, and a parent can use physical force. Their powers can contribute to the reconstruction of authority – but it might just as well lead to the very opposite: a loss of trust and respect and authorities' definitive decay.

In other words, authority is dynamic and there are different ways to (re-)convince a hesitant audience. Recognising (the potential for) resistance then becomes an important factor for its maintenance. For authority to be sustainable an actor must be able to sense and respond to another's sentiments. Such social and intellectual skills to act wisely in social relations can grouped under the heading of *social intelligence* (Salovey & Mayer, 1990: 187).

A general definition of social intelligence is the "the ability to understand and manage people" (Thorndike and Stein, 1937: 275). Others have highlighted that such skills can also be directed towards oneself. Social intelligence, by extension, then includes both

the ability to perceive one's own as well as others' "internal states, motives and behaviors, and to act toward them on the basis of that information" Salovey and Mayer (1990: 187) write. They also observe that social intelligence is often described in a more manipulative manner - as the ability to steer or manipulate the responses of others or to get others consistently and voluntarily do certain things (ibid).

Dealing with feelings and emotions, in turn, forms an important aspect of social intelligence. In the interactions with our environment, we encounter loads of affective information that must be processed. Individuals differ in the extent to which they can do so. Salovoy and Mayer (1990:189) use the concept of *emotional intelligence* to describe the "ability to monitor one's own and others' feelings and emotions, to discriminate among them and to use this information to guide one's thinking and actions" (1990: 189). Like social intelligence, emotional intelligence can thus include both knowledge about the self and about others. In essence, emotional intelligence focuses on "the recognition and use of one's own and others' emotional states to solve problems and regulate behavior" (ibid).

Earlier, I discussed how authorities use a range of "theatrical instruments" to impress their audiences (paragraph 3.4). Such expressions help them to define the situation for others – they highlight that an actor is to be trusted and respected and should be listened to. Now, it could be argued that for effectively conveying and maintaining authoritative impressions social and emotional intelligence are relevant. Especially in times of hesitance or resistance, an authority must be able to sense an audience's dismissive attitudes and convincingly respond to them accordingly. That is, one must recognise emotions in others, regulate affect and use it to motivate subjects to behave in an accepting way. But how about digital authorities? Is this not a bit too much to ask from *things*? Can digital technologies recognise, express, and utilise emotions? If we accept feelings and emotions as relevant factors in maintaining authority relationships, then investigating the relationship between digital technologies and emotions deserves a bit more attention here.

In her famous book *Affective Computing*, Rosalind Picard focuses on "computing that relates to, arises from, or deliberately influences emotions" (1997: 3). She thinks that, for computers to be truly effective, they "will have to have emotions or emotion-like mechanisms working in concert with their rule-based systems" (ibid: 12). Picard distinguishes three general emotional abilities. First, there is the ability to *recognise* emotions – "to infer an emotional state from observations of emotional expressions and through reasoning about an emotion-gaining situation" (ibid: 50). This ability can require senses such as sight, hearing for gathering facial expressions, gestures, and vocal intonation (ibid). When speaking about computing technologies there are also other relevant inputs for recognising emotions that have no equivalent in human senses such as reading infrared temperature or measuring electrodermal responses (ibid). Second, there is the ability to *express* emotions. Emotions can be relatively easily perceived (such as via facial expression, voice intonation, gestures, movement, pupillary dilation, posture), or they can

be less apparent to others and require physical contact (such respiration, heart rate, pulse, temperature, blood pressure and muscle action potential) (Picard, 1997: 27). Picard argues that computers can express emotions without having them: "the basic requirement for a computer to have the ability to express emotions is that the machine have channels of communication such as voice or image, and an ability to communicate affective information over those channels" (1997:56).[126] Last, there is the ability to *have* emotions. Picard argues that feelings are often considered to be the factor that separates humans from machines (1997: 60). Surely, the ability to have emotions is the most complex and controversial when thinking about digital technologies and it remains to be seen if it will be ever possible - and desirable - to make machines feel.[127]

Picard then extends the concept of emotional intelligence to digital technologies. Building on the three abilities of affective computing that were mentioned above, she describes a technology with emotional intelligence as one that "is skilled at understanding and expressing its own emotions, recognising emotion in others, regulating affect, and using moods and emotions to motivate adaptive behaviors" (1997:76). For the argument here, it is the recognition, regulation and use of emotions in *others* that I am most interested in. That is, when thinking about authority it does not seem so relevant if, and how, digital technologies can recognise, regulate and utilise their own emotions for themselves, but rather the extent to which they could recognise, regulate and utilise emotions of others to get things done.

The work on affective computing is interesting because it explores the potential of digital technologies to manage and utilise emotional mechanisms (Picard, 1997:84). Fact is, however, that many digital technologies (still) lack such effective emotional abilities. Not only do they not have emotions, but most computers also lack the ability to perceive and express emotions that could help them lay claim (again) to the trust and respect of a hesitant audience. For digital technologies often goes that they only have limited access to the world of social emotions: they lack emotional intelligence. Sensing the right strategy to claim authority or responding to dismissive audiences, then, is inherently problematic. Computers are designed to respond in a certain way, still generally unable to improvise and manipulate the emotions of an audience to gain compliance. When people are in doubt or unwilling to recognise digital authority, there is only so much the technology can do. It can reiterate itself or warn a user. But it generally does not have many alternatives to tailor its expressions towards them for its words to remain authoritative. When digital authority goes unrecognised, the technology needs others (designers, managers, or lawmakers as was described above) to help them out.[128]

7.6 MAKING PEOPLE THINK

In an overly complex world, people need digital authorities – for security and many other areas. Oftentimes, digital technologies "know" what they are talking about which makes their directives of great value. But as we have seen throughout this book, human beings should always remain mindful as there is no guarantee that the technology is always right. Digital technologies, for whatever reason, can be wrong and automatic compliance may have severe consequences. Moreover, authority symbols can easily be faked. Digital technologies may look good and be advertised in slick commercials but sometimes turn out to be very different in practice. In these and other ways, blind deference could hinder the very thing on which digital authority's moral justification tends to rest: improved decision-making.

The central question, then, becomes how can people know when to listen to a digital authority and when to rely on their own judgement? According to Robert Cialdini (2007: 230) saying no to a human authority begins with posing the simple question "is this authority truly an expert?" With this question Cialdini shifts our attention to the authorities' credentials and the relevance of those credentials to the topic at hand. By focusing on the *evidence* for an authority status, Cialdini concludes, major pitfalls of automatic deference can be avoided (ibid). While for digital authority this question can also be a good general starting point, determining expertise of digital technologies is often quite problematic as most people lack the general skills and experience to do so. Especially since many digital technologies are implemented with the argument of improving human decisions or are introduced as "knowledge" or "expert" systems, there is often only limited initial evidence for people to focus on.

In interactions with digital technologies, the aim perhaps should not just be to gather evidence for an initial authority status, but to maintain a focus on the authoritative communications themselves and to stimulate active assessment of what we are being told. After all, it is an old philosophical mistake – going back at least to Plato – to suppose that even with the identification of the greatest expertise, unconditional or automatic compliance is always warranted. Paraphrasing Kant (1784), the question to ask here is: how can people be urged to also think for themselves under the influence of digital authority?

7.6.1 Enhancing human efficacy

It is easy to be amazed by all incredible digital qualities that are praised as superior. But focusing on all the things a technology can do can also make people wonder: who am I to disagree? What do I know? People could feel they do not have the *capacity* to critically assess or challenge a digital technology. This brings us to the notion of *efficacy*.

Efficacy broadly refers to having the ability to achieve a desired outcome. In the context of this book this entails having the competency, skills, knowledge, and other resources to make one's own decisions. Enhancing individual efficacy would not only increase the possibility that someone recognises erroneous digital output before it is

put to use, but it can also counter the sense of powerlessness that was described above. An enhanced sense of personal or *self-efficacy* – the belief in our own abilities to reach specific goals (Bandura, 1978) - could make individuals more confident to challenge an authority as they come to view themselves as having the capacities to do so. This was already argued by Kelman and Hamilton (1989), stating that enhancing individuals' sense of personal efficacy could counteract the effects of actual and felt powerlessness vis-à-vis an authority (Kelman & Hamilton, 1989: 323).

This argument also resonates when thinking about digital authority. It was exactly this sense of self-efficacy that made Stanislav Petrov say no to the digital system telling him a missile strike was about to hit the U.S.S.R, just like confidence in their own capacities formed an incentive for the Kentucky judges who refused to rely on the risk assessment technology. An important step towards enhancing personal efficacy in human-technology interactions is to open the block-box and teach users about the workings, limits, and potential failures of digital technologies. Education and training are one way to do so.

Handbooks, presentations, and discussion sessions are examples of methods to teach individuals about the possible risks and weaknesses of digital systems. Another way to make people aware of the possibility of digital failure is to have them *experience* it first-hand. This idea originates from what is called inoculation theory (McGuire, 1961). In medicine, inoculation refers to exposing someone to a weakened form of a virus (a vaccine) that does not overwhelm the body's immune system but rather will trigger a response (the production of antibodies) that will protect against the actual disease. As described by Van der Linden and colleagues (2017), a similar logic is followed in the social-psychological theory of attitudinal inoculation: a threat is introduced by forewarning people that they may be confronted with incorrect information. Subsequently, "one or more (weakened) examples of that information are presented and directly refuted in a process called *refutational pre-emption* or *prebunking*" (Van der Linden et al, 2017:2). By pre-emptively demonstrating false claims and refuting potential counterarguments resistance is conferred. Educational methods in which individuals experience first-handedly that sometimes technologies can be wrong and that it is not just others who could be fooled by illegitimate digital authorities, can stimulate critical judgement and the potential for user resistance (see also Sagarin et al, 2002).

Familiarising people with the (in)capabilities of a technology allows them to recognise those moments in which their special attention is warranted and provides them with potential solutions for a given problem but also with *confidence* needed to act decisively. When working with a digital authority, training could leave individuals feeling better qualified to evaluate digital communications – to evaluate, question and if necessary, challenge digital output, for example by seeking support from other colleagues or initiating further investigation. As such, technology education potentially enhances personal *efficacy*.

Hence, education and training – at least to a certain extent – aims to reduce the distance between individuals and digital authority, stripping it of its mystery by revealing

what it can and cannot do. When closer to the technology, individuals are less likely to be overwhelmed by authoritative digital communications. They are reminded that, like human beings, digital authorities can be wrong. It is important to note that technology education can be general as well as specific, and ideally people living in a *digitope* are subjected to both. That is, professionals can learn about a specific digital technology and its potential benefits and risks. They are made more familiar with the technology and are provided tools for interacting with it responsibly. At the same time, education can also be dispersed more widely, to all segments of our digital society, to counter the tendency in all of us to sometimes rely on digital technologies blindly. Ideally, children in primary and secondary education are enrolled in programmes directed to making people more familiar with our digital world and the authority of things.[129] They help them to develop tools for getting by in a digital society and stimulate a healthy form of *digital skepticism*. Such digital skepticism revolves around the idea that people learn that there is always the possibility that digital output sometimes is unverified information, and that the utilisation of other available information sources is sometimes fruitful (see also Sagarin & Wood, 2007: 321). Future research could be directed towards the development of methods for stimulating such healthy digital skepticism without it becoming digital cynicism. After all, the moments we need digital technologies to make good decisions will only continue to increase.

In addition to education, ensuring enough decision-making time is also an important factor to enhance a personal sense of efficacy. In chapter 4, the notion of epistemic niche was discussed. Here, individuals are severely constrained in acquiring relevant insights from other information sources. Factors facilitating such a niche are, among others, substantial workload, and limited time availability. High workloads and little time limit the opportunity for individuals to evaluate digital output and could contribute to an experienced inequality in the relationship with the technology- feeling that they have no reasonable choice other than to comply with the technology's communications. Some studies have also found that compliance with decision aids becomes more likely when individuals' decision time is reduced (Rice & Keller, 2009; Hoff & Bashir, 2015: 416). Conversely, lower workloads could in some environments open up the epistemic niche and potentially leave individuals better equipped to come up with their own judgements and more confident to say no to a digital technology. Surely, not every decision can be delayed, and time pressure will always be part of the work of many professionals who often must decide on the spot – the parking officials from paragraph 4.4 are a case in point. In these, and all other, decision-making contexts it is important to structurally organise feedback and analyse results *ex post*. This allows those involved to reflect and learn from their own and others' decisions that and these insights could help improve future decision-making (see also section 7.7.3 below).

7.6.2 Empowering human beings

Sometimes individuals' critical judgement is impaired by the belief that they do not have a *right* to challenge authority – they feel that they are in no position to think and act for themselves. Kelman and Hamilton (1989: 322) use the term *empowerment* to refer to the process of securing the positions and powers of individuals that gives them the opportunity and the right to make their own decisions. Empowering individuals is about giving them the power to do something and stimulating them to develop an independent perspective.

Organisations and designers play an important role here. They can empower individuals by making explicit that information systems are relevant sources of information, but that there is no general obligation to follow them. An example is a facial recognition system where a possible identification - a match between a recorded image and photos in a database - is accompanied with capital letters saying:

> "This is not a positive identification. It is as an investigative lead only and it is not probable cause for arrest" (Hill, 2020).

Such messages emphasise professional roles and the agency that comes with it – that people have the right and power to act otherwise - when good reasons present themselves. In fact, organisations could even go further and make explicit it is not only the professionals' right, but their *duty* to think about what a digital system presents to them. Placing an emphasis on the responsibility of the individual – or what it means to be a "good professional" - could contribute to making people less susceptible to automatic deference to digital authority. It highlights that being a good professional does not equal blindly relying on a digital technology, just like it is not a sign of organisational disloyalty when someone seeks to critically evaluate a computer's output. While professionals who structurally and unquestionably act based on whatever a digital technology provides them might be attractive for some organisations from the point view of control (e.g., knowing users' responses and the ability to steer them) and efficiency (e.g., increased processing of cases), it does come with the risk of both individual, organisational, and societal harm. Therefore, when professionals are presented with doubtful output from a digital authority, the organisation can position itself to function as a "higher authority" and provide alternative interpretations of what a good professional is and must do. This way, the felt sense of obligation might become more directed at the organisation and its overall mission rather than to the technology itself.

Empowerment also involves providing sufficient social support for critical decision-making. Offering practical guidance, procedures and transparent structures that allow people to use their powers and challenge a technology's communications. Such structures could even be designed into the technologies – bringing us back to the notion of techno-regulation as discussed in paragraph 4.7. For example, people can be required to consult

a colleague before finalising a decision or must regularly check in with a supervisor. Implementing guidelines and mandatory moments of human evaluation into the technologies would not just remind individuals of their own powers, but also stimulate (or force) them to put these into practice when deemed necessary.

Kelman and Hamilton (1989: 322-3) explain that efficacy and empowerment are in a sense related. Empowerment can increase the skills and other resources needed for effective action and contribute to a sense of personal efficacy, while efficacy, in turn, can increase the resources needed for effective action and can increase someone's capacity to seize opportunities to use his or her powers and rights. Both empowerment and efficacy can also be linked, respectively, to the sources from which authorities (digital or human) typically derive their capacity to be consequential: the hierarchical position and coercion associated with it on the one hand, and knowledge and expertise on the other. Kelman and Hamilton argue that by means of empowerment and enhanced efficacy, individuals acquire some of the elements of authority themselves (1989: 323). Thinking about digital authority relations, both empowerment and enhanced efficacy potentially alter the relationship between human beings and technologies.

7.6.3 Intervention and compensation: handling a wrongful yes

Placing machines in the frontlines of (security) governance can have adverse effects. This can be the result of human decision-makers following a computer's output in a largely automatic and unthinking fashion and fail to recognise dubious claims even though they are faced with considerable evidence of this. But harm can also be an inevitable consequence of living and working in a *digitope* where digital authority is the result of technologies having access to a plethora of information that human beings remain barred from. Under these circumstances individual decision-makers sometimes have little rational alternative than to listen to the communications of digital technologies. Relying on digital output, then, is often the most reasonable thing one could do – even though there is no way of knowing it is right (as illustrated in paragraphs 4.4 and 4.5). And sometimes this will lead people astray and cause harm. But even though mistakes are bound to happen, this does not mean we must not seek to mitigate any negative consequences. This holds true for both individuals as well as for organisations and society at large. How do we get from saying yes, to admitting we have been wrong?

Let's start with the role of the individual's involved. While those who work with digital technologies sometimes have little choice than to accept inaccurate digital output, there is often a choice in how to do it. In many situations, professionals have at least some discretion in how to execute authoritative digital communications. As an illustration, consider some of the examples that have been discussed throughout this book – e.g. the mistaken arrests on the basis of a false facial recognition system's hit, the people who are refused to board a plane because their name is on an inaccurate digital no-fly list, police officers who shoot a

driver on the basis of a false digital claim that the vehicle was stolen, or the woman with the poodle whose car was aggressively boxed in on the highway after a computer had flagged her as potential drug trafficker. In these situations, the security officials were provided with information by a computer they could hardly ignore; they had no previous insights into the criminal backgrounds of these people and there was little or no way for them to validate the output. Yet, they were relatively free in deciding how to put that information into practice. When one can impossibly know about whether digital communications are correct, the path of action should be that of minimal potential harm in case the technology turns out to be incorrect. This means that professionals must always remain mindful about how to treat and approach those who have been flagged a security threat. With the possibility of information systems being wrong - and thus that people being stopped, searched, or arrested are innocent – the actions that follow from relying on digital authorities must always be executed as carefully and respectfully as possible. The subsequent chapter contains some additional reflections on the ethics of digital authority.

In line with the previous section, organisations and societies, in turn, must ensure there are sufficient guidelines and regulations that guide people in employing digital output – stimulating them to remain cautious when putting digital output into practice. And there are, at least, two additional responsibilities here. First, it is important to ensure that there are sufficient resources in place to discover and recover digital flaws as quickly as possible. There are many instances in which mistakes were discovered only very late or fixing things was not prioritised enough by organisations. The right responses were claimed to be difficult, for example from a technological - the case of the mistaken vehicle registration (NRC 2018) - or ownership (Nissenbaum, 1994) perspective. Either way, people ended up trapped in a bureaucratic web and unable to move forward (Widlak & Peeters, 2018). Or they remained unaware of the fact that harm was done in the first place – think of the digital errors of the Washington state prisons that were known for years before they were made known to the public.

In sum, even when the initial problems cannot be prevented, negative consequences must be limited through direct intervention and restoration of the inaccurate information in the technology and through compensation for those who became victims of such errors. This also makes that for our digital society, there must be institutions that help individuals who have been, in some way, wronged by digital authorities. The presence of independent oversight institutions and accountability mechanisms is indispensable to a just digital society.

7.7 CONCLUSION PART 3

Digital authority comes with clear limits. In the relationship with a technology, human expertise, system knowledge, institutionally ascribed authority or technological failures can hinder the experienced asymmetry that is endemic to authority. In situations, the absence of the right situational cues (e.g. design) or the disobedient responses of others can prevent people from defining something an authority situation. Sometimes there are also confrontations with multiple authority claimants, and only one gets recognised. Looking at authority's cultural layer, there is some evidence that in certain cultures people are generally less inclined to rely on digital technologies. Similarly, institutional norms, scandals, and incidents, too, could limit the construction of digital authority – especially in a context of modern reflexivity where digital authorities, too, can always be challenged and replaced. Lastly, digital authority is limited by technologies' ability to recognise, regulate, and utilse emotions to (re-)claim authority. It leaves digital authority (still) generally fragile.

Building on these limits, enhancing human efficacy and empowerment can mitigate the risks of digital authority. Moreover, individuals, organisations and societies limit negative consequences through effectively restoring inaccurate information and ensuring compensation for flawed outcomes. This is summarized in table 3 below.

Table 3. Limits of digital authority

	Relation	**Situation**	**Culture**
Characteristics	Experienced relational asymmetry	Defining an authority situation	Expectations of authority
	Consequential communications	Impressions	Conditioned values and beliefs
	Differentiated access to knowledge and force (persuasion and coercion)	Symbols (uniforms, titles, insignia, behaviors, places, etc.)	Dynamic and open to change
Digital Authority	Digital authority as asymmetrical human-technology relation	Digital technologies tapping into pre-existing authority scripts.	Narratives on authority and technology
	Digital communications to be believed or followed	Computers are social actors	Risk society and reflexivity
	Differentiated access to knowledge (epistemic niche/ opacity data and algorithms)	Situational authority cues (positions, titles, symbols, behaviors, etc.) from digital technologies	Contingency and uncertainty facilitating openness to digital authority
	Digital access to force (authorisation/ technoregulation)		
Limitation	Ascribed human authority	Absence of right authority cues	Cultural openness to innovation
	Expertise and system knowledge	Peers Subject salience	Incidents and scandals
	Technological failures and inconveniences	Conflicting (human/digital) authority claims	Reflexivity
			Organisational norms
	Social and emotional intelligence in digital technologies		
Mitigation	Efficacy (Training, education, workload, decision-making time, etc.).		
	Empowerment (Messages, practical guidance, procedures, etc.).		
	Intervention & compensation negative consequences		

8 CONCLUSIONS

We often hear that we live in a society in which authority has become problematic. It is even said that we live in a post-authority world. Sometimes the work of Hannah Arendt is cited out of its political context, and it is concluded that "authority has vanished from the modern world" (1961: 91). A look at the newspapers reveals a similar picture: parents are challenged by their children, experts are called into doubt, teachers are struggling to maintain order and there is structural violence against police officers. It seems that nowadays authority claims are questioned more often than affirmed.

But is this really true? Is there an authority crisis? While it may certainly be true that authority has disappeared from certain (political) institutions, this perhaps tells us more about those institutions than about the very idea of authority itself. Narratives of a demise or erosion of authority in modern society are often misplaced and tend to overlook authority in its variety and ubiquity in so many human activities. We simply cannot do without authorities - nor do we typically want to. Our lives are permeated with authorities that are searched for and willingly followed because they allow us to deal with a great deal of complexity. We cannot know everything ourselves, and in many instances – both personally and professionally - we rely on authorities to function properly. With this book I suggest that authority did not simply disappear from our world but is rather in a state of shift. To understand our modern world and the actions of human beings in it, we must search for new places of authority – beyond its traditional institutions, positions, and actors. This search has led us to the domain of technology.

Digital technologies are involved in an increasing number of human decisions. And in more and more situations, we witness a great willingness to comply with all sorts of digital output. Some of my favorite examples involve GPS systems leading drivers astray. But notwithstanding the potential dangers of drivers' blind reliance on these technologies, in a rapidly digitising world the potential impact of this phenomenon is much larger. In this book I considered the field of security in which digital technologies now play a role in dealing with problems that significantly affect individual lives. Technologies now advise and instruct about things such as imprisonment and release, about arrests and further scrutiny, about life and death in military operations, about access and restriction and about

parking tickets and fines. The potential impact of such interventions on individuals and societies means that there is a sense of urgency to this book.

Despite sustained efforts of philosophers of technology the role and influence of technological artifacts is still often ill-understood in many contexts. By discussing their authoritative potential, I highlight that digital technologies can be much more than instruments or assistants for improved decision-making - as they are often depicted in popular discourse. The goal of my philosophical exploration here was to develop the idea of digital authority. This concept offers a framework for reflecting on the compelling influence of digital technologies. These are much needed to develop effective and ethical solutions for the serious problems that can arise from uncritical reliance on digital technologies.

This chapter provides an overview of the ideas and findings of this book. Subsequently, I offer avenues for further thought and research. This chapter ends with a glance into the future.

8.1 THE RESEARCH DOMAIN: AUTHORITY, TECHNOLOGY AND SECURITY

This central question of this dissertation is: To what extant can digital technologies be perceived as authorities? While this question is not limited to a particular domain as we now interact with digital technologies pretty much everywhere, the field of security has been my main scope given its centrality and potential impact on human life. To answer this question, the first part of this book explored the research domain. In chapter 2 and three the terms digital security governance and authority were discussed.

Over the last decades, our world has become saturated with digital technologies. I described our world as more and more resembling that of a digitope. It constitutes an environment in which capable, connected and information generating technologies are now everywhere around us. This new environment increasingly supplants that of nature and refines the technological system, or technotope, as described by Jacques Ellul and others.130 We can already hardly escape the web of digital technologies that surrounds us. It increasingly facilitates life and shapes how we think and act. Digital technologies rapidly become a prerequisite for our survival.

Against this background, I discussed in chapter 2 how digital technologies are central in many contemporary governance structures. I used the term digital security governance to refer to the application of a broad range of digital technologies to collect, process, store, edit and distribute the general information for security decision-making (Eijkman, 2014). In the field of security, we see police officers check a digital database if a vehicle is stolen, airport officials are warned by a computer about “high-risk” passengers, judges are digitally advised about parole decisions, expert systems help child protection services to better protect the well-being of children and fraud is now often flagged by digital technologies.

This book is about decision-making contexts in which decisions result from the interplay between humans and digital technologies – in contrast to fully automated or algorithmic decisions where human intervention is taken completely out of the loop. I am fascinated by all these emerging situations in which machines provide users with information, advice, or instructions on how to think or act. The aim is not to replace humans here, but rather to help them make better choices.

Digital technologies are often depicted as an instrument to improve human cognitive capacities, as a complementary means to better cope with everyday complexities and uncertainties. However, as we have observed, decision-making is not necessarily enhanced with the introduction of digital technologies and digital output can become much more than just a complementary source of information. In fact, as several scholars have highlighted, what the computer says can become a heuristic that limits vigilant information seeking and processing. This would not be problematic when technological thinking is always right, however, unfortunately this is not the case. I have discussed several anomalies in the use of digital technologies that could lead people astray. These can result from problems in the digital technologies themselves (examples of such *technology-based anomalies* are inaccurate data and bad algorithms) as well as from people using the technology in the wrong way (examples of such *user-based anomalies* are mistakes caused by a lack of attention or general skills). And while most of us know that digital technologies are not flawless, there are still many situations in which we end up blindly relying on them – even when the stakes are high and when contradictory information is readily available to us. The real issue then, as the famous North American psychologist Burrhus Skinner once said, becomes thus not whether machines think, but whether people do (Skinner, 1969: 288).

In the third chapter of this dissertation, I discussed the general idea of authority. The word authority is derived from the Latin auctoritas, a word used in many ways, but which usually describes the ability to exert a compelling force on people "greater than that of simple advice, but less than that of command" (Mommsen, 1887 c.f. Arendt 1961: 123). Building on this discussion and the ideas of Bruce Lincoln (1994) in particular, I have come to consider authority as the ability to produce consequential communications. That is, communications that motivate an audience to respond in a way that signifies acceptance. It is a special kind of influence, distinct from coercion and persuasion alike. With authority, one does not need to argue, threaten, or force someone to bring about conformity. Rather, people accept the communications of authorities out of trust and respect or do so out of a strategic willingness to pretend this were the case (Lincoln, 1994: 9). Authority rests with the voluntary recognition of an audience and is always open to contestation and revision. But once recognised, people tend to do or think that whatever is communicated to them is worthy of their unquestioning acceptance, regardless of their own judgement. We then seem to think and act in the world through someone else.

In this book I described authority from an interactionist stance, as emerging from the exchanges between an actor and his audience within a given setting. Someone (or something) does not simply have authority. It is not intrinsically linked to individual qualities such as office, wealth, ancestry, or charisma. Rather it is a result, or the outcome of social interactions. To expand upon this idea, I discerned three conceptual layers. First, I described authority as a relational concept. Authority implies an asymmetrical relationship in which one is recognised by the other(s) not as an equal but, in some sense, as (a) superior. It is this posited, perceived or ascribed relational asymmetry, Lincoln writes, on which the consequentiality of authoritative communications rests as it permits certain actors "to command not just the attention, but the confidence, respect, and trust of their audience - or (..) to make audiences act as if this were so" (Lincoln, 1994: 4). Second, I identified authority's situational layer. The experience of authority relations is intrinsically grounded in the character of the situation. As illustrated by a wide range of psychological research, authority situations can incite accepting, reverent and submissive responses. There is whole a range of gestures, settings, and items that convey authority and invite recognition of an audience. That is, they leave an impression on an audience that makes them define the situation as an authority situation, leading them to follow, accept or obey whatever is communicated to them. Third, I discussed the cultural layer of authority. Many of our ideas and responses to authority are culturally shaped. Throughout our lives, within the family, at school and later in public and professional life, we learn who to trust and respect. This way our ideas and expectations of authority are conditioned and guide us in recognising and responding to authority.

In sum, this layered perspective on authority emphasises that authority is constructed in relations, situations, and cultures. It avoids treating authority, as often happens in discussions, as a static or stable entity but instead highlights its dynamics and open-ended essence – how authority claims and their recognition may conflict and change over time. As such, it also allows for reflections on potential new authorities.

I then introduced the idea of *digital authority* which can be defined as the capacity of digital technologies 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. The term also refers to those technologies that are commonly considered to have such a capacity. It presupposes a world where authority is no longer a purely social phenomenon – only to be found in relations between human beings. The following paragraphs condense the findings of this dissertation and further explain the potential construction of digital authority as well as its limits.

8.2 UNDERSTANDING THE AUTHORITY OF THINGS

Following the discussions on digital security governance, the changing context of human decision-making and the meaning of authority, the second part of this book was dedicated to answering the defined main question. Chapters 4, 5 and 6 each investigated a distinct aspect of digital authority. In chapter 4, I focused on the relationships between humans and technologies. I discussed various human-technology relationships and suggested that authority relations with technologies do not seem to fit into existing categories. After having introduced the idea of digital authority relations, this chapter further investigated its characteristics and constructive elements. In the fifth chapter I examined digital authority situations. I used Erving Goffman's interactionist notion of "definition of the situation" to describe how actors may come to define for others what is going on and trigger corresponding responses. In this chapter I discussed the cues and underlying mechanisms that can turn human interactions with a technology into recognised authority situations. In the sixth chapter I shifted attention to the cultural context in which attitudes and expectations towards authority and digital technologies are conditioned. How are digital technologies viewed and portrayed in modern-day societies? And is there a place for digital authority in our modern world?

8.2.1 Asymmetries in Human-Technology Relations

Don Ihde described how people can relate to technologies in different ways. In chapter 4, I described that authority relations with technologies, however, have their own distinct characteristics. While in authority relations human perceptions and actions are actively shaped by a technology, these relations go beyond the existing structures of technological mediation that Ihde discerned, as the technology becomes the terminus of our experience. The technology is explicitly recognised as an other based, similar to what what Ihde termed alterity relations (Ihde, 1990: 97-108). Alterity and authority relations are both grounded in the experience that we are interacting with something other than ourselves. This makes that not all technological artifacts are potential authorities. We will not recognise a pair of glasses, a pen, or an axe as authorities because they are not a source of interaction. But with digital technologies we can experience a sense of otherness, due to their apparent autonomy and the interactions they give rise to. More specifically, with their digitisation, technologies have begun to communicate. And it is this capacity of producing communications that draw our attention and trigger responses, that allows for thinking about authority relations with things. For a long time, authority was deemed a purely social relationship, as it were only people who issued communications to be believed or followed. But now with the advent of digital technology, the communications of things are also actively recognised by people – whether it was a bleep, an artificial voice, or a text on a screen. That is, they have become an explicit source of communication and interaction. Yet, authority relations with technologies go beyond simple alterity as well. Digital authority relations

are recognised as fundamentally asymmetrical and come with a belief that a technology has a right to be listened to. It is the experience of this asymmetry that motivates people to do or think whatever a technology communicates. In an authority relation, the influence that technologies exert on human perceptions and behavior is so compelling that it is characteristic for the relationship between human beings and their world. We do and think things simply because the technology tells us. One could say that in these configurations we experience the world by a technology. But how to understand the emergence of this asymmetrical human-technology relationship in the first place? While answering this question covered the entire book, chapter 4 continued to discuss some initial elements that I consider as constitutive parts of digital authority relations.

The asymmetry that is endemic to digital authority relations can be linked to ideas differentiated access to knowledge and power. First, much like human epistemic authorities (such as teachers, scholars, and other experts), digital technologies can be recognised as having access to a world of relevant knowledge from which others are largely barred. Especially in environments with limited alternative sources of information (epistemic niches), individuals may come to depend on whatever is presented to them by a digital technology. I used modern-day parking enforcement as a case in point. But even when additional information is available, people can still easily consider the technology as having superior access to large volumes of data and complex algorithms, underlying the digital output that is presented to them. The inherent opacity of digital systems can lead people to presume that compelling reasons exist for the thing someone is asked to do or think. I linked this discussion to the modality of persuasion that is constitutive of many authority relationships (as discussed in paragraph 3.3). While persuasive technology is a fascinating topic in its own right (Fogg, 2003), with digital authority, I argued, it is not so much about actual acts of persuasion but rather about the experienced possibility thereof – its presence being recognised in the background of the relationship. Rather it is the belief that for any digital utterance, precisely the right fact, statistic or reference can be put forward from that volume of data that persuades. Surely, most digital technologies are not programmed to explain themselves, and generally must rely on others for justifying their output – which, in turn, people may not understand very well. But as long as people think that logical evidence for digital claims exist somewhere, there is no need for persuasion in the first place. Hence, the asymmetry of digital authority relations, ultimately, rests with the presumption that persuasive arguments can be brought forward upon request. It is this presumption that motivates people to trust and respect the technology and potentially go to great lengths in accepting their communications. It leaves people persuaded in advance.

Second, in their relations with digital technologies people can also recognise inequalities in access to power or coercive force. As is the case with persuasion, authorities must not force or coerce someone into doing something. But when present in the background, in a state of latency as Lincoln has put it (1994: 6), the coercive capacities of

an actor can co-shape the asymmetry that is characteristic of authority relations. However, digital technologies generally do not hold formal positions and powers that allow them to make and enforce decisions. They seem unable to command obedience in similar ways as human authority figures, such as judges, police officers, and military officials can. Yet, there is an increasing number of situations in which professionals are formally required to consult digital systems for decision-making– think of the judges making parole decisions or customs officers at airports – and compliance could potentially be enforced by managers or other human authority figures. But technological access to coercive force is generally conditional upon the presence of rules and human authority. But upon closer investigation the potential of coercion in human-technology relationships can also rest more directly in the technology itself. I turned to the domain of techno-regulation and how non-confirming behavior can be designed out – think of computerised access gates that only open for certain individuals or forms that will not save unless all cells are filled in. Techno-regulation offers opportunities to reflect on the coercive potential of digital technologies. It highlights how such powers become increasingly embedded in the technology itself and that others are no longer needed to execute them. Compliance, then, could be secured with the alternative of force lurking in the background of the relation. While someone may disagree with the technology, he still might choose to do as he is told because he knows that the technology will otherwise coerce him into obedience – the gate will not open, the document will not submit, or the system will not respond. In fact, as technologies are not emotional beings feeling sorry for their subjects, the idea of coercion lurking in the background can be extra powerful – technologies do not make exceptions.

This way, experienced differences in access to knowledge and power are tied to the commodities of persuasion and coercion present in the background of the relation. In latent form, they are constructive of the asymmetry of authority relations with technologies, producing attitudes of acceptance, trust, respect, or the willingness to act is if this were so. In practice, technologies often still must rely on others (humans) to provide persuasive arguments or use coercive force in times of resistance – as further discussed in chapter 7. However, when force or persuasion becomes necessary there is no digital authority, it is just force or persuasion. Rather, it is about the assumption that "the computer knows" or that it can legitimately force, that ensures accepting or compliant responses in advance.

8.2.2 Digital Authority Situations: resembling the social

In chapter 5 I explored the possibility of authority situations with digital technologies. When people enter a situation, they ask themselves: What is going on? And what responses are appropriate here? (Goffman, 1986; Van den Berg, 2009). To answer these questions, people rely on learnt scripts. Scripts are packages of knowledge that allow people to navigate their everyday life. They specify the roles of actors and objects within a given situation and contain certain action patterns. An evoking context, with the right cues in

the environment, will trigger specific scripts. There can be signs present in the physical environment or in the actions of others that will lead us to interpret and respond to the situation in a particular way. Scripts are thus anchored in a specific social setting and help an individual to infer their role and how to move forward.

The notions of situations and scripts are also important for understanding how people come to recognise and respond to authority. Scholars such as Arendt (1963) and Milgram (1974) already argued decades ago that authority lies vested in the character of the situation. I have explained how settings with the right cues trigger scripts that lead individuals to define something as an authority situation and incite them to take up accepting, reverent or even submissive roles. Authoritative items, places, symbols, or demeanors are situational cues that help individuals recognise what is going on and trigger scripts that offer a mental shortcut for choosing the appropriate responses. In some way, those claiming authority by relying on these cues end up defining the situation for us – their performance effectively can turn a situation into an authority situation when the situational cues are recognised as such by others present in that situation.

But what does this mean for the idea of digital authority? After all, previous discussions all focused on situations in which people interact with other people and entered an authority script. The remainder of the chapter was dedicated to explaining how digital technologies, by tapping into pre-existing social scripts, can create authority situations for us as well. Digital technologies, I began my situational argument, can call forth many social responses in human beings - the type of responses we traditionally reserve for interactions with other humans. I discussed two strands of literature: *anthropomorphism* in which humans' explicit responses to digital technologies are central and the *Computers Are Social Actors* (CASA) paradigm that focuses on our mindless responses. Both lead to a clear and consistent conclusion: digital technologies give rise to a range of very familiar social responses by human beings, and for this the social cues given off by the technology do not have to be very sophisticated or obvious, nor do people have to be aware of their social responses to technologies.

Perhaps there is no need any more to make strong distinctions between human and technological authority situations: digital technologies can become central in authority situations in very similar ways as human authorities do. They can mimic traditional notions of authority and as such trigger human beings to respond in accordance. By relying on the same array of props and impressions as human authorities, digital technologies can come to shape situations for us. For example, digital technologies now fill roles that were previously fulfilled by human authority figures, such as doctors, teachers, experts or police and security officials. These roles suggest authority. There are given all sorts of titles that further emphasise this - technologies that are labelled "expert" or "specialist" or are referred to as digital teachers and doctors. And sometimes there are more sophisticated cues, think of a particular language style or visual representations of authority symbols such

as certificates or logos. All these cues that play a role in the construction of human authority situations can also evoke people's recognition when exhibited by digital technologies.

In short, digital technologies may not look or act in the exact same way as humans to lead subjects to recognise authority situations and respond to it accordingly. And it does not matter if their knowledge or powers are real - whether technologies are legitimate authorities or not. What matters is the authoritative impression they can make and that can trigger people to enter an authority script. In these situations, an important addition, digital technologies do not function as a proxy for human authority. Several experiments found that individuals are not directed at those behind the computer (the programmers or their bosses) when responding in a social manner. Rather, linking back to the earlier discussion on human-technology relations, it is the digital technology to whom we are oriented. This way, with the right situational cues, authority situations with digital technologies can emerge.

8.2.3 Cultural openness to digital authority

Expectations and ideas of authority relations and situations are culturally conditioned. In chapter 6 culture was described as the "software of the mind" (Hofstede et al. 2010: 5) – it shapes the mental program that guides much of our thinking, feeling, and acting throughout our life. Cultural psychologist Geert Hofstede explains that culture is a collective phenomenon, covering many social notions and rules that are learnt and shared with others who live or lived within the same environment (Hofstede et al, 2010: 6). When thinking about authority, cultures differ in the extent people are socialised into following a hierarchy and recognising inequality in social relationships. Some cultures are more authority oriented than others – dictating a strong respect for traditional authority figures. Could it be that only in these cultural environments digital authority claims are likely to resonate? I do not think so. Throughout this book I have argued that authority is an indispensable part of human life. Across cultures and societies authorities are willingly searched for to get things done. Cultural differences towards authority, therefore, are primarily a reminder that to yield authority effectively, actors must tailor their performances to a specific setting. At most, I concluded, it tells us that there are different ways technologies could successfully claim authority – for example through performances that extrapolate hierarchies or that are more egalitarian in essence.

Similarly, people are socialised in different technological environments. Some societies are generally characterised by a more positive outlook to digital technologies and a greater willingness to rely on them than others. However, concerns about innovative technologies are widespread and we have seen examples from people's blind reliance on a technology from across the globe. I therefore decided to zoom in on how people are directed towards digital authority within societies. In essence, in many modern cultures there is a continuous oscillating between two conflicting narratives on technology – of accepting and rejecting, of embracing and distrusting, of portraying technology as either good or bad

in public discourse and popular culture. People learn about how tech will iron out every difficulty and fix most of our pressing problems (Morozov, 2013; Broussard, 2018), while they are simultaneously familiarised with the risks that come with it. The field of security is no exception here - digital technologies are often simultaneously depicted as a valuable solution to crime and other security threats as well as a potential source of risk and threat themselves. People in modern cultures are often confronted with these fundamentally different narratives. What these narratives have in common, however, is a deterministic take on the influence of technologies on societies – technologies are believed to either fix or ruin our world. By accentuating their shaping or transformative capacities, technologies become something potentially awe-inspiring. They can foster deferent attitudes towards digital technologies and facilitate the construction of digital authority relationships. At the same time, being in a society with strong positive and negative technological ideas also can lead people to interpret and respond to technologies in contradictory ways. Generally, it is an environment of ambivalence that can both stimulate and limit digital authority.

For a more encompassing understanding of the cultural foundations of digital authority, I turned to human experience of the modern world itself. Modernity is characterised by people's increased awareness of uncertainty and complexity of their world. Much of this has to do with the erosion of tradition and religion and threats that are now increasingly linked to human activities, leaving people with an increased sense of worry and responsibility for what is to come. Initially, the growing importance of science and technology in modern societies provided people with the impression of more certainty and control over the world. However, with science being an inherently skeptical endeavor, where claims to knowledge are constantly challenged and revised (Giddens, 1998: 23-4) and when the severity of technological catastrophes became apparent, it ultimately led to a range of new uncertainties and anxieties. A risk culture or society developed in which there is a general climate of uncertainty and an obsession with future harm that could befall us.

Under conditions of modernity, people are increasingly occupied with creating and revising their certainties. To make choices and to cope with uncertainty, individuals reflexively monitor their world. They are now constantly critically assessing what is going on and actively weighing up who to trust or rely on. In a society after religion and tradition, in which risks are experienced as human responsibilities, individuals must seek new relevant expertise and reliable others to listen to. Paradoxically, awareness of the limitations of expert knowledges and institutions and their susceptibility to revision and change has made room for other authorities to emerge. Individuals become more open to new authorities that could them control their world and counter feelings of anxiety and doubt Most individuals will navigate modern life by relying on a series of different authorities. And with digital technologies being presented as effective solutions against uncertainty and complexity, as described above, modern societies provide fertile environments for the construction of digital authority.

8.3 DIGITAL AUTHORITY, ITS LIMITS, AND IMPLICATIONS

In the introductory chapter, it was stated that computers incite us to rethink many common beliefs about the nature of artifacts and people. I hope to have also, perhaps authoritatively, challenged the assumption that it is people who make things listen and do. As it turns out, technologies are also increasingly capable of grabbing our attention and securing our compliance as well. It urges us, among other things, to reconsider our conceptualisation of human beings as "users", as in many situations we seem no longer seem to primarily "use" the technology, but rather strongly respect and follow it.

In our world, the concept of authority has a lot to offer when seeking to understand people's willingness to go great lengths under the influence of digital technologies. By distinguishing the different layers of authority, the authoritative potential of digital technologies becomes apparent. In each layer, relevant developments can be discerned that now allow for the construction of digital authority. The emergence of digital technologies has given rise to new, interactive human-technology relations that are recognised as fundamentally asymmetrical. Situations with digital technologies can be experienced as authority situations and trigger accepting and compliant responses of individuals. And in modern cultures, people are increasingly open to new authorities that could help them knowing and controlling their uncertain world. I have come to view digital authority as the outcome of the conjuncture of an authority relation, situation and culturally conditioned expectations and ideas. When these combine, interactions with a digital technology may produce attitudes such as trust, respect, acceptance, or obedience among human beings. Authority, then, is no longer exclusively human.

Still, as indicated in chapter 7, digital authority has clear limitations and there are many things that obstruct the construction of digital authority. Technologies do not hold the formal offices and powers that are constitutive of many human authority figures. There are also still plenty of areas in which individuals find themselves more knowledgeable or authorised to rely on their own judgements and make their own decisions. And surely digital authority is not exempted from the very reflexive process of which it is the product.

But above all, digital authority is still limited by technologies' capacity to observe and effectively respond to critical moments of potential crisis. These are the moments in which people, explicitly or implicitly, say no to a technology. Technologies lack a mind and do not have intentionality or autonomy as human beings do, leaving them with only a limited action repertoire to deal with resistance and challenge. They only respond the way they are programmed leaving them often only able to reiterate their output to an unwilling human subject. Most of the time, technological persuasion or coercion to gain compliance is only latently present. Digital technologies generally have to rely on others to win over their audiences again – for example for marking out the potential benefits of compliance or by highlighting the negative consequences of their behavior. Their absence makes deviations

from digital communications more likely. This makes digital authority generally fragile and constrained for now.

With digital innovation becoming increasingly synonymous with the continued development of Artificial Intelligence this can all easily change in the next decades. The identified limitations will likely turn out to be current limitations. Digital technologies will more and more convincingly present themselves as authorities and become increasingly resilient towards acts of resistance. But the first digital authorities have already presented themselves in today's world. Both in our personal lives and in multiple areas in public life we can already see the potential of digital authority. While it provides us with the opportunity to make informed decisions that we would have otherwise not been able to make, it has also generated new incidents and risks.

A common thread in this book has been the stories in which people clearly go too far in following digital authorities. Examples from the field of security illustrate the potential impact on human lives. With the security domain becoming increasingly technological, and with more decisions resulting from human-technology interactions, the effects of wrong or mistaken digital output will only continue to grow. Through enhancing individual efficacy and empowerment some of the potential pitfalls of digital authority can be mitigated. For example, education, training, and clear guidelines could stimulate and enable individual judgement and decision-making. Last, it is important to recognise that bad things can always in complex decision-making environments, including those in which we sometimes simply have no rational alternative than to rely on digital technologies. Individuals, organisations and societies must therefore remain mindful in how to employ the digital output in their daily operations as well as build institutions that help individuals that have been wronged by digital authorities.

8.4 ONGOING CHALLENGES AND FUTURE DIRECTIONS

This book provides a philosophical framework for understanding digital technologies and their potential to compellingly influence human beings. It allows for novelle interpretations of existing research and opens paths for future scientific studies. It also offers insights for the mitigation of associated risks and the design of effective solutions. As an epilogue, I present some avenues for further research and the future development of the digital authority paradigm.

8.4.1 Empirics and experiments

Much of this book draws on anecdotal cases, existing studies, and philosophical analysis. Real life examples were mainly gathered from desk research. Over the last years, I also interviewed experts and visited organisations in the field of security – such as the police, tax authority and (local) government – to learn more about the digital technologisation of

modern society and the field of security. It provided me with practical examples of digital authority as well as with invaluable insights in the status quo in the field of security and its complexities. Among other things, it has taught me that despite the many impressive modern technological developments and Hollywood blockbusters, we are far from living in a world that is ruled by complex Artificial Intelligence.

My personal life has also been a constant source of real-world examples over the last years – as many authors will recognise. I remember the moment I opened our living room curtains while looking at my phone to check the weather app whether it was raining. Or the argument that unfolded on a Saturday morning with my son in the backseat of the car hysterically claiming it was night because the car LCD screen was (incorrectly) showing a night sky navigation map.

I consider such experiences and empirical examples important to illustrate and conceptualise the idea of digital authority. A next step now would be to supplement these conceptualisations with more case study research. Detailed insights into practices of digital security governance are important to shine more light on how humans and digital technologies interact in real life. Indeed, there are several common limitations within the field of security – such as difficulties in gaining information and access to relevant sources - but findings of studies into the use of risk assessment tools (Dekkers, Van der Woude & Koulish, 2019) and predictive policing systems (Meijer, Lorenz, Wessels, 2021) have already clearly indicated the value of this demanding type of research.

Another approach would be to conduct several experimental studies. In this book we have already seen the importance of such research. In chapter 5, for example, ample attention was paid to the CASA experiments. Some of those, such as the emergency robot discussed in paragraph 5.6, already gave us some interesting insights into how, and under what conditions, people were willing to follow a digital authority. Similarly, the research of B.J. Fogg (2003) at Stanford University on Persuasive technology, studies into automation bias (Skitka, Mosier and Burdick, 1999) and experiments on trust in autonomous machines (Waytz, Heafner and Epley, 2014) have highlighted the value of experimental research designs. Designing experiments would allow to test a set of basic questions and assumptions about digital authority further in a lab-controlled setting. How do peers influence the willingness to recognise digital authority? When exactly does technological failure result in loss of authority? To what extent do the stakes of a decision– such as in the field of security – determine deference to a technology? What are effective trainings to enhance human efficacy in digital situations? etc. The study of authority has a long history in experimental research, and many questions relevant for understanding digital authority have been investigated before in a social context. As such, these studies offer a relevant starting point for digital authority experiments. However, as we have learnt from past authority experiments, experimenting with authority, in any form, is not without ethical controversy. Designing and doing experiments on digital authority should therefore be done with care.

8.4.2 Ethics of digital authority

How to deal with ethical issues of digital authority? Even when technologies are designed and implemented with good intentions (such as helping people to make better choices or enhancing efficiency and effectiveness in decision-making processes), ethical concerns may arise. And when technologies become recognised as legitimate authorities, there are several moral questions. Digital authorities, for example, can make people do immoral things, threaten independent human judgement, or have undesirable societal consequences. While this book has predominantly focused on the conceptualisation of digital authority, addressing the ethical aspects of digital authority would make a logical next step.

Ethics and authority have a complicated relationship. Authorities obviously help individuals to deal with many things and are indispensable for people and societies to function. As such, a willingness to follow instructions from those who know – whether human or technological – is generally a good thing. But it can become problematic when recognised authorities lack ethical convictions themselves. Hannah Arendt (1963) described how the many horrors of history are traceable to the human willingness to do whatever authority figures say. In the same spirit, Kelman and Hamilton (1989) described the My Lai Massacre as a crime of obedience. Their observations were echoed by infamous studies like Milgram's shock experiments that further highlighted authority's potentially evil outcomes.

The relationship between ethics and technology is equally complex. Technologies come with ethical praise as well as with condemnation (see chapter 6). There are many technologies that have clearly made lives easier, safer, healthier, or more comfortable. However, other technologies are evaluated more negatively. Coal plants, for instance, are now viewed as a significant environmental threat. Nuclear weapons are considered a frightening source for suffering and destruction. And even an emancipatory technology as the birth control pill is still contested in conservative religious communities because it would violate the "natural course of things" as argued elsewhere (Verbeek, 2011: 3).

Ethical discussions about technology and authority both run the risk of taking the form of an encompassing critique. Classical approaches to the ethics of technology typically considered the overall dangers posed to the human world by the phenomenon of "Technology". Similarly, discussions about authority have long been overshadowed by a general discomfort of everything that has to do with the concept. These approaches are not particularly helpful as they remain distant from ethical problems related to actual technologies in their context of use.

Existing ethical frameworks for technologies would offer a good starting point for the moral reflection on digital authority. For example, there are relevant frameworks for evaluating the ethical aspects of persuasive technologies (Berdichewsky & Neuenschwander, 1999; Verbeek, 2006b). Although, unlike persuasive technologies, digital authorities are not necessarily intentionally designed as such, these frameworks help to identify relevant points of application for moral reflection. They direct our attention to the intentions of

the designers and whether these can be morally justified, the actual performance of the technology and on how it influences the experiences and actions of human beings in a use context. For the design of ethical digital authorities these are all highly relevant. They make visible both the intended and unintended consequences of a technology and could guide moral assessment among all those involved. Thinking critically about the practice of digital authority could also help design barriers to prevent harm and establish accountability. This constitutes a constant process as not all outcomes can always be foreseen – things are frequently more about anticipation than prediction (Albrechtslund, 2007: 63).

Further attention to the ethics of digital authority is valid and vital. But the aim would not be to prevent the construction of digital authority. While digital authority can have ethically undesirable consequences, the opposite is also true. That is, legitimate digital authorities often help us to do things right and spur social progress. Not having ethical digital technologies, then, would also be unethical. A broad philosophical investigation, as also stated by Luciano Floridi (2013: xii-iii) on the ethical impact of digital technologies on human life and society is therefore essential if we want to anticipate the difficulties and identify opportunities of our expanding digital world.

8.4.3 Accountability and blame

Accountability in a digital society is about people taking ownership of technological mistakes or malfunctions that have caused harm to others. But despite its importance in a world where people are exposed to new digital harms and risks, it seems that accountability is systematically undermined and becomes increasingly difficult in situations where technologies become perceived as authorities.

Decades ago, in her article *Computing and Accountability* (1994), Helen Nissenbaum already identified several barriers to accountability for computerised failures, risk, and harm. For example, digital mishaps are often the work of "many hands". As an illustration, Nissenbaum used the famous case of the radiation treatment machine *Therac-25* which caused massive overdoses to several patients. It took many months of study to discover that there was not a single source for the origin of the malfunction. There were numerous faults such as software errors, failing hardware, bad interface design, inadequate testing and quality assurance, overestimations of the system's reliability and inadequate hospital investigation (Nissenbaum, 1994: 76). Leveson and Turner (1993) analysed the Therac-25 case and concluded that it is very difficult to blame (only) the person operating the machine as many hands contributed to the harmful outcome. Earlier, John Ladd made a similar argument stating that with bad outcomes that follow from people's interactions with computers there is often "a long distance between the inputs of particular individuals and their significant outcomes, and there are so many different kinds of input from so many different sources (designers, operators, managers, and consumers), that as far as establishing human responsibility is concerned, we find ourselves adrift in a vast sea of

anonymity where responsibility has become so diffused as to evaporate into nothingness". (Ladd, 1989: 209).

Nissenbaum (1994: 77) also addressed the idea of the computer as scapegoat as a barrier that obscures accountability for digitally mediated harms and risks. For different psychological reasons we are quite comfortable with saying that it was "the computer's fault" or a "computer error" (Nissenbaum, 1994; Friedman, 1995; Moon & Nass, 1998; Turkle, 2005). Blaming a computer allows people to shirk responsibility, avoid potential repercussions or save face in interactions with others (Kalman, 2015).[131] But, as Nissenbaum says, sometimes people are just genuinely perplexed when they point towards a computer – it is then a response to circumstances in which accountability has become so blurred that the computer indeed becomes the most salient actor (Nissenbaum 1994: 78). The computer serves as a replacement for an absent other who is, or should be, accountable.

The fact that we can easily transfer responsibility to a computer not just facilitates digital authority but also further obscures accountability. In authority relations, people tend to no longer see themselves as responsible for their own actions - instead responsibility is attributed to what is considered a legitimate authority. At this point, people no longer view themselves acting in a morally accountable way, but rather as an agent for an external authority (Milgram, 1974: 8). In the case of digital authority people can come to think of themselves as intermediaries between the world and the technology that is telling them what to do or think. For them, blame has shifted to the technology, further eroding a sense of accountability and responsibility within their roles.

Decades after Nissenbaum's article, the need to rescue accountability has only become more pressing in a world in which technologies have come to play a larger and more influential role in human affairs.[132] Individuals increasingly find themselves in a context where decisions are no longer made by identifiable rulers and bosses but are now the result of black-box systems and organisations. Dan Davies (2024) describes in his book *The Unaccountability Machine* that modern organisations are often designed to reduce accountability and make it problematic to assign blame when things go wrong (Davies, 2024: 4). It requires us to structurally evaluate and adjust ideas of accountability to fit today's complex socio-technological reality.

When accountability is absent, there is no answer for digital harm and risks. They become tragic incidents or unfortunate consequences of technological innovation. But where the standard of accountability is upheld, Nissenbaum already concluded, there is a foundation for just punishment and compensation of victims (Nissenbaum, 1994: 73). It offers a general environment that encourages earnest and responsible practices. This makes discussing and upholding accountability in a highly technological world with digital authorities emerging and making things ever more complex a challenge well worth pursuing.

8.5 THE FUTURE - WHERE IS THIS GOING?

This book sketches the contours of an emerging world in which authority applies to both people and things. I am curious to see how this will develop in the next decades. Although it is impossible to predict what our digital future will look like, I am certain that the computer will continue to compellingly say no - and yes. And it will do so at more places and at fascinating levels of autonomy. Smart cars, facial recognition surveillance systems, automated profiling techniques, semi-autonomous weapons and robotics are still much in their infancy. Our world, including the field of security, will continue to develop as one of algorithms, data, and Artificial Intelligence. And somewhere ahead of us we might even find technologies in more formal positions of authority, being granted legitimate powers and rights over people. Hollywood has already envisioned its extremes with numerous films about AI and robots ruling the world.

We can never know what the future holds. But maybe the computer can? Hoping for an authoritative answer I asked an AI-chatbot about the future of "computer says no":

> "The future of digital authority will likely involve a continued shift towards greater use of technology and data in decision-making, as well as ongoing debates about the appropriate balance between the benefits and risks of this trend".[133]

Makes perfect sense.

8

NOTES

1. INTRODUCTION

1. There is also the story of a man who wanted to see the Pope but ended up a thousand miles away in Germany after typing 'Rom' instead of Rome into his car navigation system (The Daily Mail, 2019).
2. As a case in point, see also Follow the Money (2023).
3. The term was coined by Eijkman (2014:116) to refer to "the collection, processing, storage, sharing and application of digital data for risk profiling". I use a broader definition that also allows us to include other issues beyond risk profiling such as the use of digital systems and data for detecting and handling violations (such as in traffic or fraud).
4. Hawk-eye is a digital system for score detection. While it is correct most of the time, there can be flaws. Take, for example, a disputed call in the 2007 Wimbledon men's final between Federer and Nadal (from Collins & Evans, 2008: 293). A ball was hit by Nadal which appeared to land well beyond the baseline. While both Federer, the audience, and the umpire all seemingly agreed, Hawk-eye called it in, and the umpire accepted this digital judgement. The situation was described by The Telegraph (2007):

   > "Federer, a tennis conservative, has always been against the introduction of Hawk-Eye, and he was as angry as he had ever been on Centre Court when an "out" call on one of Nadal's shots was successfully challenged by the Spaniard in the fourth set. The HawkEye replay suggested that the ball had hit the baseline; Federer thought otherwise. It was then that Federer asked umpire Carlos Ramos whether the machine could be turned off. Ramos declined but also seemed to suggest that he had thought the ball had landed long. The Hawk-Eye review gave Nadal a breakpoint, which he converted for a 3–0 lead, and Federer continued to complain during the change of ends. "How in the world was that ball in? S***! Look at the score now. It's killing me, Hawk-Eye is killing me," the Swiss said. So, a system which was introduced to prevent McEnroe-style rants at officialdom actually left one of the sport's gentlest champions fuming."

   Also, in football there has been controversy, for example in 2020 when Hawk-eye failed to register a goal in the Premier League's game between Sheffield United and Aston Villa. It turned out that all seven cameras used by the systems were obscured. The incident was heavily ridiculed on social media and the operator of the goal line technology later apologised for the error. The game ended in a 0-0 draw (BBC, 2020). For an academic reflection on this technology see Collins & Evans (2008).
5. A *decision* refers to the choice to do something or to behave in a certain way (Szaniawski, 1980: 328). Behavior, in turn, is understood here in its broadest sense, including the externally observable activities of the individual (e.g., the execution of a particular activity or the abstention from action) as well as the adoption of purely mental attitudes (e.g., acceptance of a particular view) (ibid.). A decision always

implies that there are alternatives. A choice to behave in a certain way always presupposes at least two options: to behave that way or not. Decision-making is the process of making this choice between alternatives.

6. Simon proposed the term 'bounded rationality' to emphasise the limitations of humans' capacity to reason and choose (1972:162; 1997: 93-4).
7. According to Simon, people tend to *satisfice* rather than aiming for the best decision; they reduce cognitive effort by making decisions that are good enough, which allows them to stop exploring alternatives. In his words: "The Scottish word satisficing (= satisfying) has been revived to denote problem solving and decision making that sets an aspiration level, searches until an alternative is found that is satisfactory by the aspiration level criterion, and selects that alternative" (Simon, 1972: 168).
8. Mosier and Skitka (1996) call this tendency to use automated cues as a heuristic replacement "automation bias".
9. AT5 (2018), Apple-Navigate stuurt fietsende toeristen door de Piet Heintunnel'.
10. Ihde (1990:141) refers to this transforming capacity of technology as "technological intentionality". He writes: "Technologies, by providing a framework for action (..) form intentionalities and inclinations within which use patterns take dominant shape" (ibid.). Rather than neutral instruments, technologies *invite* certain types of interaction – as put elsewhere they have, as it were, an 'intention' with their users (Achterhuis, 1997: 153 c.f. Van den Berg, 2009: 31).
11. Chapter 4 also includes a discussion on the concepts of *affordances* (Norman, 2013; Gaver, 1991) and *nudges* (Thaler and Sunstein (2008).
12. In Migram's initial experiments, a subject entered a laboratory with the idea of participating in a study of memory and learning. He or she was assigned the role of a teacher and is asked to teach word associations to someone. The learner was seated in another room in a miniature electric chair with his arm strapped against excessive movement and an electrode attached to his wrist. He was administered electric shocks of increasing intensity by the learner whenever he made an error. The teacher was seated before a shock generator with a lot of switches labelled with a voltage designation ranging from 15 to 450 volts (Milgram, 1973: 62). The teacher was given a sample shock of 45 volt to strengthen his or her belief of the authenticity of the machine. While each teacher was a genuine naive subject, the learner was in reality an actor who will not be shocked at all. The aim of the experiment, Milgram writes, was "to see how far a person will proceed in a concrete and measurable situation ìn which he is ordered to inflict increasing pain on a protesting victim" (ibid.).
13. In the years that followed, many others built on this research and aimed to discover more about the nature of human obedience and compliance (Bock & Warren, 1972; Kilham & Mann, 1974; Burley & McGuiness, 1977; Meeus & Raaijmakers, 1995; Blass, 1999; Burger, 2009). Later this paradigm was also transferred to the domain of human-computer interaction to discover if people were willing to torture a robot on the command of an authority figure (Rosalia et al, 2005; Bartneck & Hu, 2008).
14. Conversely, inanimate objects can also not be subject to authority, De George explains:

> "We own things; we have dominion over them; we may have the authority to use them or to dispose of them. But they constitute, in these instances, the field over which authority extends, not those subject to authority. The case of animals is more ambiguous. We give animals orders, and well-trained ones respond appropriately. Whether we wish to call such responses authority responses and whether we wish to say that animals are subject to authority depends on one's view of animals as well as on one's characterization of authority" (De George, 1985: 16).

Nonetheless, in the context of this book we focus on human beings as subject to authority and follow De George to the extent that all kinds of authority are characterised by human social contexts.

15. Woods (2009), for example, speaks of the authority of systems, meaning that "the automated system is capable of taking over control of the monitored process from another agent if it decides that intervention is warranted based on its perception of the situation and its internal criteria" (Woods, 2009: 6). While the capability to intervene autonomously can facilitate the kind of authority I am interested in, it only becomes relevant in relation to human beings.
16. As will be explained in chapter 3, this definition is rooted in the work of Bruce Lincoln (1994).
17. According to Abrahm Maslow's (1943) *Hierarchy of Needs*, safety and security are basic needs, to be placed directly after physiological needs (food, water, warmth, and rest).
18. A frequently made distinction - that is also encapsulated in the etymological roots of security, which is derived from the Latin compound *sine cure*, meaning "without care" or "carefree"- is between security in objective terms (i.e., the condition of not being threatened) and security in subjective terms (i.e., the condition of being untroubled by anxiety or feeling secure) (Zedner, 2003:155).
19. Surely, the implementation of new security measures and technologies does not happen in a vacuum. For example, new legislation and regulation and large-scale terrorist attacks such as 9/11 have been important drivers behind this development.
20. In the case of technological innovation, it is often more appropriate to speak of *perceptions* of what people will do with a technology. As explained by Frissen and van Lieshout (2006), there are often quite significant contradictions or discrepancies between the expected or perceived user needs and behaviors in the early stages of technological innovation and actual social practices when these innovations find their place in daily life. As they illustrate the example of Short Message Services (SMS):

    > "The huge success of Short Message Services (SMS), particularly among young people, came largely as a surprise to the industry, and thus cannot be understood by looking at the way user needs and behavior were perceived in the design and marketing stages of the ICT application. In these stages, the 'ideal user' of SMS was imagined as a businessman who used SMS rationally and instrumentally for time-saving and planning purposes. Among young people, however, SMS became extremely popular for continuous connectivity with their peers and for creating their own symbolic environment. And this in spite of the design of the interface and service, which are not particularly suitable for these purposes, at least at first sight." (Frissen & van Lieshout, 2006: 253-4).

21. See for example, Wolff (1990).
22. Adorno identified authoritarian submission (towards ingroup authority figures) and authoritarian aggression (against those who violate conventional rules) as two personality traits that would make a person susceptible to totalitarian or anti-democratic ideas. While heavily criticised for bias and methodology – and Adorno even distanced himself from the empirical research (c.f. Rood, 2013: 107) - the book has been hugely influential.

2. DECISIONS IN A DIGITOPE

23. In his essay *Intelligent Machinery* (1948), Turing described a machine that consisted of:

    > "An infinite memory capacity obtained in the form of an infinite tape marked out into squares, on each of which a symbol could be printed. At any moment there is one symbol in the machine; it is called the scanned symbol. The machine can alter the scanned symbol, and its behavior is in part determined by that symbol, but the symbols on the tape elsewhere do not affect the behavior of the machine. However, the tape can be moved back and forth through the machine, this being one of the elementary operations of the machine (Turing 1948: 3).

24. The ENIAC, one of the first 'supercomputers' that was developed in the 1940's, weighed more than 27,000 tons and occupied approximately 167m2.
25. The goal of this section is not to provide a complete overview of the field of Artificial Intelligence. For a more in-depth discussion on this issue see for example Russel and Norvig (2021) or Zhang & Lu (2021).
26. Others speak of 'strong' or 'full' AI (Searle, 1980). There is still much discussion on the definitions and various forms of AI.
27. Also labelled 'weak' or 'narrow' AI (see also note 26 above).
28. Lo is an archaic term used to draw attention to an interesting event.
29. Pew research Center (2019) even reported that nine-in-ten Americans use the internet.
30. For more insights on who's not online (and why) see the survey study of Zickuhr (2013).
31. The concepts of ambient intelligence and ubiquitous computing are also related to this section (Riva et al., 2005; Brey 2006; Bick & Kummer, 2008; Van den Berg, 2009). Bick & Kummer describe that both "characterize intelligent, pervasive and unobtrusive computer systems embedded into human environments, tailored to the individual's context-aware needs. Such miniaturised modern information and communication technology (ICT) supports humans by offering information and guidance in various application areas" (Bick & Kummer, 2008: 79).
32. Epistemologically there is a difference between data, information and knowledge. Information is made up of a collection of data and knowledge is made up of various threads of information. I will, however, follow Cukier (2010:2) here and use "data" and "information" interchangeably because it is increasingly difficult to distinguish the two. As Cukier writes "Given enough raw data, today's algorithms and powerful computers can reveal new insights that would previously have remained hidden".
33. Hilbert estimated that in 2007 stored digital data doubled every three years (2012: 9).
34. Another recent estimation is that now more than 16 zettabytes of data is generated annually and that this will have grown tenfold by 2025 (Reinsel, Gantz & Rydning, 2017).
35. This example is derived from Schonberger and Cukier (2014: 9). They used the great Library of Alexandria and calculated that "the digital deluge now (..) is the equivalent of giving every person living on Earth today 320 times as much information as is estimated to have been stored in the Library of Alexandria".
36. Today, televisions, thermostats, cars, home appliances and many other products are increasingly embedded with processors, software and connected to the internet, enabling them to capture and share data. Our smartphones contain GPS chips to determine location, microphones to measure sound, cameras to capture images and several other sensors (Susskind, 2018: 49). Other, devices use sensors to measure things such as temperature, pressure, noise, pollution, humidity or the proximity of another object (e.g., in self-driving cars), to detect motion, recognise people or to scan number plates or barcodes. Sensors translate physical phenomena into digital data and as such play an important role in the *datafication* of our society (Broeders et al., 2016: 42).
37. Broadly speaking, the term governance refers to patterns of rule. It encompasses the intentional activities of a particular actor to shape the flow of events or the "business of managing our world" as others have put it (Wood & Shearing, 2007: 6). In public administration the term governance often has a more specific connotation, associated with a more horizontal and multi-actor mode of steering (post)modern society (Prins, 2014: 69). Mark Bevir (2009) distinguishes between old and new governance. The latter is rooted in the work of theorists such as Thomas Hobbes and Max Weber and views state government as the dominant actor, capable of effectively steering society in a hierarchical and technocratic manner. With new governance, by contrast, state government is no longer so superior in the organisation and management of society (Bevir, 2009). Cooperation with a partiality of non-actors has become increasingly important for state governments to achieve its goals (Anttiroiko et al, 2011: 2). Hence, in the context of this book it is therefore also important to note that I focus on both governments and private organisations in the governance of security.
38. Williamson (2015: 90) speaks in this context of a shift from "governance with a voice" to "governance with a brain".

39. For a more detailed discussion on the different types and levels of human interaction with digital technologies see Parasuraman et al. (2000).
40. Bovens & Zouridis exclusively focus on public executive agencies.
41. Snellen (1998) and Zuurmond (1998), too, see a fundamental change in the way these organisations function. They see that today's professionals - they specifically speak of street-level bureaucrats - are controlled through infocratic means, instead of through bureaucratic (or organisational) structures. In their view the use of ICTs facilitates higher levels of managerial control, resulting, among others, in a reduced scope of action for professionals.
42. As Salinger describes it: "bounded rationality means that individuals (..) act purposefully, but not necessarily as if they are both fully informed and perfectly rational" (2010: 71).
43. See note 7 above.
44. For example, one study created a smart mutant mouse called "Doogie". This genetically engineered transgenic mouse showed enhanced learning and memory as well as greater ability in learning new patters (for the technical story on the manipulation of the NMDA (N-methyl-D-aspartate) receptor's subunit in the mouse's brain see Tang et al., 1999).
45. TMS is a magnetic method used to stimulate parts of the brian. Studies have showed, among other things, that TMS of the motor cortex can improve performance in a procedural learning task and various other areas (for a more detailed discussion see Sandberg & Bostrom, 2006a: 206).
46. Sandberg and Bostrom (2006b: 2-4) further divide internal cognitive enhancement technologies into whether they are hardware of software. On the one hand, internal hardware deals with biological modifications such as genetic modifications, nutritional interventions, surgery, neural implants, or pharmacological cognitive enhancements. Internal software, on the other hand, aims to enhance cognition by learning improved cognitive strategies, for example via education, mental training, or therapy.
47. Because anomaly has a more neutral connotation, I prefer to use this term over words like error, problem, failure or fault. Moreover, the word anomaly has a clear starting point the expectations of the individual (or a society) about the outcome of the use of a digital technology. Errors, on the other hand, are "occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency" (Reason, 1990:9). However, as will be discussed in this paragraph, digital errors are just one of the reasons why people might end up doing the wrong thing.
48. Note that this category can overlap with the 'user-based anomalies' as explained below. That is, inaccurate data and flawed digital output can be a result from both the involvement of others in designing or using the system, as well as from the actions of users themselves.
49. I am aware that bad quality of video images can stem from both software and hardware issues. It affirms the point made later this paragraph that the categories I suggest here are artificial and can easily overlap.
50. See also the story of a Michigan man who was arrested because of a faulty facial recognition match documented by Hill (2020).
51. This category corresponds to Reason's later discussion – based on Rasmussen's (1986) performance levels - of errors that relate to *Skill-based level* performance (Reason, 1990:56).

3. AUTHORITY

52. Augustus. Res Gestae Divi Augusti, 34.2. For further discussion of the role of the Res Gestae in justifying Augustus' rule, see Bosworth (1999).
53. We could complicate things further through the distinction between *de facto* and *de jure* authority. Like the idea of executive authority, *de jure* authority implies having an official right according to some formal system of rules and procedures. Such authorities derive their official status from some formal

framework, set of rules or custom which provides them with the right to exercise power (Kekes, 2003: 77). My interest, however, lies primarily in the more practical idea of *de facto* authority. Nevertheless, it is important to note that *de facto* and *de jure* authority are not opposites: *de facto* authorities may, or may not, be also *de jure* authorities, and *vice versa* (Kekes, 2003:77).

54. Also remember the distinction between authority and coercion in the Latin *auctoritas*, where it was typically opposed to the possession and use of *potestas* (power or force).

55. See Friedman's discussion on the surrender of private or individual judgement, in which "the subject does not make his obedience conditional on his own personal examination and evaluation of the thing he is being asked to do" (Friedman, 1990:64).

56. Stanley Milgram also refers to this idea when he speaks of the agentic state which he defines as "the condition a person is in when he sees himself as an agent for carrying out another person's wishes" (1974: 133). In this state of agency, the individual no longer thinks and acts for himself but experiences himself as an instrument. He uses this term in contrast to that of autonomy – when a person sees himself as acting on his own.

57. Academic scholars and religious figures, for example, are often depicted as having a unique relationship to science and God. They are posited – spiritually and epistemologically - as intermediaries between the lay community and the thing on which these others depend for their knowledge or expertise.

58. Goffman defines performance as "all the activity of a given participant on a given occasion which serves to influence in any way any of the other participants" (Goffman, 1959: 26).

59. At the same time Goffman notes that there also signs and expressions that may be emitted unwittingly and unwantedly. These signs may be seen as characteristic for that person (Goffman, 1959: 14). Hence, performances do not consist of the impressions an actor *gives*, but also consists of the signs they *give-off*. Even when a performer stops his show, he may still be giving-off expressions because the audience will continue to focus on the signs that are there for them to be seen.

60. This again also ties in with Goffman's discussion on the various items that make up a performer's 'personal front' (Goffman, 1959: 34-6).

61. Patrick Wilson (1983) argues in his book, *Second-hand Knowledge: An Inquiry into Cognitive Authority*, that we are inclined to recognise others as authorities, at least in first instance, when those of whom we ourselves think well of do so as well. This is the omnipresent phenomenon of personal recommendation.

62. Pierre Bourdieu's uses the concept of *habitus*, to describe how this process of cultural learning as society being "writing into the body" (Bourdieu, 1990a: 63). Habitus is a tacit system of dispositions that influences how we experience and act in the world. The habitus is acquired through internalisation of social conditions, opportunities, and restraints of the environment in which live. As Bourdieu writes:

> The conditionings associated with a particular class of conditions of existence produce *habitus*, systems of durable, transferable dispositions, structured structures predisposed to function as structuring structures, that is, as principles which generate and organize practices and representations that can be objectively adapted to their outcomes without presupposing a conscious aiming at ends or an express mastery of the operations necessary in order to attain them. Objectively 'regulated' and 'regular' without being in any way the product of obedience to rules, they can be collectively orchestrated without being the product of the organising action of a conductor (Bourdieu, 1990b: 53).

63. This view aligns with some prominent ideas on the related concept of power. Both Michael Foucault and Bruno Latour already wrote about the interactive character of power in detail. We find in Foucault's argument that power as such does not exist in people or institutions but is fluid and part of relations and every interaction (1977: 32-57). Latour, in turn, claimed that "power is not something one can possess - indeed it must be treated as a consequence rather than as a cause of action" (1984: 264). Power, he continues, "may be used as a convenient way to *summarise* the consequence of a collective action (..) it may be used as an effect, but never as a cause" (1986: 265).

4. RELATION

64. This is not an isolated example, limited to the United States. Recently, there was a similar case in the Netherlands. A Dutch prisoner was released 401 days early due to a "human error in combination with limitations of the information system" (Algemeen Dagblad, 2024). Once the mistake was discovered, the man who was released prematurely was found to be missing.

65. The second reference to "human" in this schematization is intentionally lowercase and emphasises the accepting or compliant role of the human subject in a digital authority relationship.

66. After fierce public protests the ban was lifted just 19 days later (Marshall, 2014: 362).

67. In Europe, a system called *disc parking* was introduced. Disk-parking allows drivers to park for free for a given period of time, as long as a disc was displayed indicating the time that the vehicle was parked. It was first introduced in Paris in the 1950s in order to regulate on street-parking and has been adopted in various other countries since (Ison & Budd, 2016: 152). While this system is not flawless as it can be easily manipulated by dishonest drivers – it too allowed for more effective parking enforcement as officials could now concentrate themselves on the disc behind the windshield rather than purely on their own observations and memory.

68. Sometimes the dispatching of parking enforcement officers on scooters is done automatically (NOS, 2017).

69. Or take the example of Dutch police officers tasked with fighting drug-traffickers described by Jan-Kees Schakel and colleagues (2013). For humans it is largely impossible to discriminate drug-trafficking behavior in large traffic flows: there is simply too much going on and many indicators are only assessable after a vehicle is stopped (Schakel et al, 2013: 179). Hence, a digital system was developed for detecting criminals trafficking heroine and other drugs via highways. The system focused on those vehicles traveling to and from the cities of Rotterdam and Maastricht within a given short period of time – a typical drug-trafficking route from one of the world's largest harbors to a city providing access to the European hinterland. In addition, a database was used with license plates of vehicles that were seen near coffee shops in Maastricht (ibid). Strategically placed cameras along the busy route (with between 3000 and 4000 vehicles passing every hour) automatically read license plates of the vehicles driving by. For those that met the criteria of the profile, the system generated an alert to which police officers could respond. The results were impressive. Several drug-traffickers, carrying kilos of soft and hard-drugs were caught. Even a taxi that normally would not have been stopped, was now inspected and a significant load of hard drugs was found. Unfortunately, there were also some painful misses. For instance, an ambulance that was transferring a patient from Maastricht academic hospital to the hospital in Rotterdam was wrongly pulled over, as was an elderly lady who was flagged as a drugs-runner by the system when she was driving her small poodle dog from the doctor's office. Multiple police cars boxed her in, forcing her to stop on the highway after which she was subjected to an unpleasant search. Later it was concluded that she had done nothing wrong (Rienks, 2015: 136). However, as the technological capabilities far exceeded human capacities to process (or gather) all the information needed for decision-making here, unquestioning acceptance of the system's output seems a sensible thing to do.

70. Hardwig uses the formula: "B has good reasons to believe that A has good reasons to believe that p" (1985: 338).

71. When there is no other information available to base one's decision on, one could say that it is not so much about *believing* a particular digital claim as true but more about *accepting* the claim (Cohen, 1992). In the case of parking enforcement, a parking officer may believe that a car is not illegally parked because he has seen the driver buy a ticket, and still have digitally generated information that the car is parked illegally. He may accept the digital output and write a fine on the grounds that the system's judgement is more reliable than his own.

72. Van den Hoven (1998:101) refers to this as the 'pressure condition'.

73. This observation raises new moral questions. Can we, for example, blame a person relying on a technology

in an epistemic niche? After all, the police officer, the customs agent, the military official, or the bank clerk, etc. are often *casually* responsible for the damage done - they are the ones making the final critical decisions. However, the question of *moral* responsibility and who can be held rightly responsible for events where digital technologies are involved, turns out to be much more problematic (Snapper, 1985; Ladd;1989; Van den Hoven, 1998; Rooksby, 2009). Van den Hoven has argued, that in decision-making contexts where individuals have no rational alternative than to do or think what a digital technology presents them on a screen, they cannot be deemed responsible for unfortunate outcomes as they cannot control their own thoughts and actions (Van den Hoven, 1998). Others have nuanced this view by emphasising that those working with digital technologies often have at least some discretions in how to employ the digital output that is presented to them (Rooksby, 2009). Even when it makes perfect sense to rely on digital authority, as long as alternative courses of action are available, a person can still be held responsible for harm being done. See also paragraph 8.4 for a discussion on the ethics of digital authority and the issue of accountability.

74. In this context John Danaher (2016) speaks of the threat of *algocracy*: the "situation in which algorithm-based systems structure and constrain the opportunities for human participation in, and comprehension of, public decision-making (Danaher, 2016: 245) and the opacity problem (the potential incomprehensibility of digital output to human reasoning).

75. As Rooksby argues, we are in no way forced in systematically believing the information generated by such systems, just like we are not forced to believe, let's say, the speedometer of my car or the thermometer outside of my window. While I often assume, with good reasons, that these technologies are accurate, it is "hardly a matter of psychological necessity that they do so" (2009: 84). There is always the freedom to do or believe something else.

76. This quote was, among others, used by John Stuart Mill in an article on the subject of pledges (1832).

77. However, also the legal status of digital technologies is a constant topic of debate. In a recent case described about an Air Canada chatbot that gave incorrect information to a traveler, the airline pointed at its chatbot and argued it is "responsible for its own actions" (BBC, 2024). While the chatbot had promised a discount that was not available to a passenger, Air Canada argued that the chatbot was a "separate legal entity that is responsible for its own actions". However, a court ruled that Air Canada was responsible and that "it makes no difference whether information comes from a static page or a chatbot". Companies remain liable for what the tech says and does.

78. This also makes incorporating techno-regulation a very cost-effective way of regulating. By designing the enforcement of a rule into a digital technology, expensive law enforcement personnel is less needed (Van den Berg & Keymolen, 2017: 191).

79. With techno-regulation the forces that steer, guide and influence behaviors also frequently go unnoticed by its subjects. As Van den Berg and Keymolen highlight:

> "One key characteristic of techno-regulation is that end users are often entirely unaware of the fact that that their actions are being regulated in the first place. Techno-regulation invokes such implicit, almost automatic responses, that end users do not realise that their action space is limited by the artefacts' offerings" (Van den Berg & Keymolen, 2017: 191).

Especially when a techno-rule goes unnoticed, it becomes very difficult – or simply impossible – to disobey this rule they conclude. Artifacts and systems, then, become implicit managers and enforcers of rules: "they automatically, and oftentimes even unconsciously, steer or guide users' actions in specific directions – toward preferred (set of) actions and away from undesirable or inappropriate ones" (ibid.).

5. SITUATION

80. Two years later, U.S. Senator Ted Stevens stated in a committee hearing that his wife Catherine was also subjected to similar questioning at an airport about whether she was Cat Stevens (The New York Times, 2006).

81. See for example Sharkey (2008) or Noflylistkids (2024).

82. This was, among others, mentioned in a conversation I had with one of the parents whose child was on the no-fly list. To learn more about experiences with no-fly lists, I had an online call of an hour with a representative of the Canadian NGO 'No Fly List Kids' in January 2019. The organization was founded by parents of children falsely flagged on Canada's No Fly List. As listed on their website, their aim is to "ensure that the character rights of all Canadians, including those wrongly affected by the PPP [Passenger Protect Program], are protected" (Noflylistkids, 2024).

83. Goffman admits that there are some issues with defining and thinking about situations. Among other things, situations generally cannot be reduced to monolithic proportions. He writes:

> "It is obvious that in most "situations" many different things are happening simultaneously – things that are likely to have begun at different moments and may terminate dissynchronously. To ask the question "What is *it* that's going on here?" biases matters in the direction of unitary exposition and simplicity." (Goffman, 1986: 9).

Moreover, he continues, that speaking of *current* situations can also be misleading as it can easily foster the impression that there is a clear demarcation in time and space, while "the amount of time covered by "current" (just as the amount of space covered by "here" obviously can vary greatly from one occasion to the next and from one participant to another" (ibid.).

But despite these unintended connotations that come with the notion of 'definition of the situation', it has become indispensable to the interactionist understanding of role performance in relation to specific situations making it impossible to ignore or replace it as some have stated (Van den Berg, 2009: 191). While being aware of these limitations and the warnings that come with it, I will follow Goffman's example in this case and "will let sleeping sentences lie" (Goffman, 1986: 10).

84. Furthermore, given the context of this book, note that the idea of scripts is also used by researchers in various ways when thinking about technology. For example, scholars in Science and Technology Studies (S&TS), may speak of scripts to describe the vision of users and use practices that are inscribed in technological artifacts by designers (Akrich, 1992: 208). Others have also used the notion of scripts to argue how technological artifacts themselves can function as scripts in everyday environments (van den Berg, 2009: 193-219). My interest in this chapter, however, lies more in the human episteme in relation to practices of (inter)action.

85. The media equation is the product of a research program known as the *Computers are Social Actors* (CASA) paradigm. Within this program a wealth of experimental evidence has been gathered for people's tendency to treat computers as real people. While it is beyond the scope of this book to provide an all-encompassing literature review (e.g., see Reeves & Nass, 1996; Nass & Moon, 2000; Johnson & Gardner, 2009), there are four broad areas of CASA-experiments: social rules and norms, social categories and roles, personalities and attraction and emotions. Taken together they reveal that many of the classical insights from social science on human-human interaction also hold for human-computer interaction.

86. Other CASA research on social rules and norms also revealed that computers can flatter (Fogg & Nass, 1997); that people use social rules regarding praise and criticism in their interactions with computers (Nass & Steuer, 1993; Nass, Steuer, Henriksen & Dryer, 1994; Nass, Steuer, Tauber & Reeder, 1993); and that we also apply the common norm of reciprocity when working with a computer (Fogg & Nass, 1997; Moon, 2000).

87. A follow-up study also found that female-voiced computers were perceived to be more socially attractive and trustworthy and that subjects identified more with computers that matched their own gender (Lee, Nass, Brave: 2000). Others experiments on the use of social categories and roles showed that the "ethnicity" of a computer has a profound effect on one's attitudes and behaviors (Ibster & Lee, 2000); that computers can be teammates (Nass, Fogg & Moon, 1996) and scapegoats (Moon & Nass, 1998); and that labelling a technology a specialist leads people to appreciate it more (Nass, Reeves & Leshner, 1996) - see also paragraph 5.6.

88. Note that in practice, mindlessness and mindfulness are both matters of degree - the degree of active information processing can vary within and between individuals (see also Sauer et al, 2013).

89. This is in line with what Masahiro Mori (1970) suggested earlier. In his famous article *The uncanny valley* he argues that when humans are interacting with humanoid objects (things that appear nearly human) and are suddenly reminded that they are not real, the natural expectations of their interaction are broken resulting in strange and uncanny feelings. Mori gives the example of unexpectedly shaking a prosthetic hand. When people shake a prosthetic hand but expected a natural hand because of its human-like appearance (having veins, muscles, tendons, fingernails, fingerprints, and colors resembling human pigmentation), they might be surprised by the cold temperature or the lack of soft tissue (1970: 33-4). In these kinds of situations, a sense of strangeness arises; it is uncanny.

90. To these types of cues, Nass (2004: 37) has added that social responses can also be encouraged via the manifestation of emotions, the presentation of faces, the (perceived) engagement with and attention to the user, autonomy and through unpredictability. These elements are all typically possessed by humans, and therefore, when computers, robots or other technologies exhibit one or more of these characteristics they seem likely to evoke mindless social responses.

    In *Persuasive Technology. Using Technology to Change what We Think and Do* (2003), Fogg also distinguishes five general categories of social cues that evoke individuals to draw inferences about the social character of a computer: physical cues (the presence of physical attributes such as a face, eyes, body or humanlike movement), psychological cues (including the display of emotions, personality, motives, preferences and the use of humor), language cues (the use of (interactive) spoken or written language and the ability to recognise a user's speech), social dynamics (following the tacit rules associated with interpersonal interaction such as turn taking, praising, responding to questions and cooperation) and lastly, occupying social roles such as teammate, pet, tutor or opponent.

    There is substantial overlap between the categories of Fogg (2003) and Nass (2004) and colleagues (Nass & Steuer, 1993) and there is a general agreement about two fundamental points Johnson & Gardner (2009) write. First, there is agreement that computers do display cues that suggest humanness. Second, scholars agree that these cues cause individuals in a mindless state to behave socially (Johnson & Gardner, 2009:152).

6. CULTURE

91. The case was extensively described by Widlak & Peeters (2018).

92. Tokmetzis (2012) has described this case in detail.

93. The Fiscal Information and Investigation Service (FIOD) is an agency of the government of the Netherlands responsible for investigating financial crimes.

94. Cited from Hofstede et al. (2010: 88). Original article "Verkoper eist legitimatie van koning" in NRC Handelsblad, 23 December 1988.

95. The term Latin Europe is used to refer to countries such as France, Italy, Portugal, and Spain. They share a common history as they were all a central part of the Roman empire. They speak languages derived from the empire's common language – Latin. Moreover, they are all predominantly Catholic countries which has shaped many of many of their values and attitudes (Pérez-Perdomo & Friedman, 2003: 1).

96. As also described by Malcolm Gladwell (2008: 236-42) in his bestselling book *Outliers*.
97. Hofstede does not argue that certain scores, on any of his dimensions, are better or worse. He also does not claim that a culture's characteristics are representative of the behavior of every individual – people have their distinct personalities and someone from a high-power distance culture can still hold more egalitarian views.
98. Elsewhere, Hofstede refers to the comparative studies that show that French companies have one or two hierarchical levels more than comparable companies in Germany and the UK. French managers also generally hold more privileges, get paid substantially more and are less accessible than their German colleagues (Hofstede, 2010: 31- 40). As he also points out, the normalisation of social inequality is also expressed in how top managers of big companies are addressed. CEOs in France, for example, are called Mr. PDG – a prestigious abbreviation meaning President Director General (Hofstede, 2024).
99. For an alternative study, see also the Network Readiness Index (NRI) that was first published in 2002 by the World Economic Forum, Cornell University and INSEAD. In 2022, the NRI covered 131 nations and is now a publication of the Portulans Institute, focusing on the readiness to use and development of ICT in different countries. See Dutta & Lanvin (2022).
100. This example is taken from Februari (2023: 26-27).
101. There is even a popular criminological hypothesis stating that suspects are more likely to be acquitted when technology fails to reveal sufficient forensic evidence –it is termed the "CSI-Effect" (Davis et al, 2010).
102. Late 2023, an agreement was reached within the European Union on the Artificial Intelligence Act. As a response to existing worries, this regulation "aims to ensure fundamental rights, democracy, the rule of law and environmental sustainability are protected from high-risk AI" (European Parliament, 2013). It is the first comprehensive law on AI by a major regulator.
103. Eidinow (2007: 139-155), for example, describes how in the classical Greek and Roman societies, people used curse tablets to try and influence the gods. These were typically very thin sheets of lead with a curse inscribed and used by people to ask the gods to intervene –think of curses that call down vengeance on someone who committed petty theft or love spells for intimate personal relationships.
104. Bernstein (1996) links risk to the early Italian *risicare* meaning 'to dare' (Bernstein, 1996: 8).
105. In his book *Liquid Modernity* (2012), Bauman argued that "flexibility has replaced solidty as the ideal condition to be pursued of things and affairs" (Bauman, 2012: ix). Elsewhere he elaborates:

> "Forms of modern life may differ in quite a few respects – but what unites them all is precisely their fragility, temporariness, vulnerability and inclination to constant change. To 'be modern' means to modernize – compulsively, obsessively; not so much just 'to be', let alone to keep its identity intact, but forever 'becoming', avoiding completion, staying underdefined. Each new structure which replaces the previous one as soon as it is declared old-fashioned and past its use-by date is only another momentary settlement – acknowledged as temporary and 'until further notice'. Being always, at any stage and at all times, 'post-something' is also an undetachable feature of modernity. (..) What I've chosen to call (..) 'liquid modernity', is the growing conviction that change is *the only* permanence, and uncertainty *the only* certainty." (Bauman, 2012: viii).

106. Giddens also draws our attention to the accessibility of expert skills and information to lay actors (Giddens, 1991: 30-2). He describes that in traditional society expert knowledge was largely unavailable to lay individuals, for example because of illiteracy (Giddens, 1991: 30). Nowadays such knowledge is much wider available – especially with the expansion of the internet. Rather, the question has become whether one has the skills to correctly interpret and work with such knowledge. Anyone can become an expert in the era of late modernity. As put elsewhere, "true experts who have the proper education and skills are forced to spend time on self-improvement and to fight against "unreal" experts" which, in turn, even creates greater risks, mistrust, skepticism (Volkova and Pruel, 2019: 53).

107. In chapter 7 I will discuss how modern reflexivity and technological risks complicates the durable construction of digital authority. But let's conclude for now that contemporary cultures seem to be increasingly open to new and digital forms of authority.

7. HUMAN SAYS NO

108. The alarms sounded during a period in the Cold War in which the Soviet-American relations were severely strained. The United States President, Ronald Reagan, had only recently called the Soviet Union "an evil empire" and explicitly rejected calls for a nuclear freeze (The New York Times, 1983). In turn, the Soviet leader in the early 1980s, Yuri Andropov, expressed his worry of an American attack after had Reagan announced plans for a European missile defense system. About this Andropov said that Washington was searching for "the best way of unleashing nuclear war" (c.f. Garthoff, 1994: 111). The extent of the tensions was already highlighted three weeks earlier when a South Korean Boeing 747 en route from New York was shot down by a military jet after it had entered Soviet airspace, killing all 269 people on board. NATO responded with a show of military exercises. Against this background, the threat of nuclear missiles was immediate and vivid for those in Serpukhov-15.

109. It was estimated that a "counterforce strike" in which the Soviets and United States only attacked military targets, could have killed more than 50 million people in short term (Daugherty, Levi & von Hippel, 1986; Levi, von Hippel & Daugherty, 1987). But things would probably have gotten much worse (Bracken, 2012: 84-90). In 1983, a few months before the alarms suddenly sounded, a war simulation on the beginning of a nuclear exchange with the Soviet Union was conducted in the United States. Despite that the participants were instructed to follow the current military doctrines (i.e., to respond proportionality, firing only at military targets rather than counter value in which high population area are targeted), it was found that none of them managed to hold back in their response. Every participant escalated and began to fire missiles at high population areas. The game calculated a short-term death toll of at least half a billion and another half a billion, due to fallout, starvation and ongoing war in the months that followed.

110. See, for example, Economist (2017).

111. After the alarms fell silent, his colleagues praised Pretrov for his decision to trust his own judgement. However, an official investigation resulted in a reprimand for the lieutenant Colonel. He received an official warning for failing to correctly fill in his logbook. "Because I had a phone in one hand and the intercom in the other, and I don't have a third hand," he told The Washington Post (1999). The story of Stanislav Petrov remained a secret until 1998, until the incident was described in the memories of his chief, general Yuri Vontintsev. Petrov was then honored by the United Nations, received the World Citizen Award and the Dresden Peace Prize. He also starred in a documentary. Petrov received hundreds of letters of recognition from all over the world and several Hollywood stars insisted on meeting him. But while for many he became a "hero of humanity", Petrov never thought of himself as such.

112. A relevant example is also the study of Dekkers, Van der Woude & Koulish (2019). In their study, they describe how security officials use the smart camera system 'Amigo-boras' for migration control purposes. The authors found that the security officials using this vehicle to check vehicles often operated independently of the technology. Most of their decisions were taken based purely on their personal assessment, and much of the system's output was ignored. About those vehicles that were stopped, and which were flagged by Amigo-boras, the officers stated that they would have spotted the vehicles also without the technology (Dekkers et al., 2019: 245-6).

113. Blass (1995; 1996) and Altemeyer (1981; 1988) shared similar conclusions.

114. For a more complete overview of the role of personality in the Milgram obedience experiment see Blass (1991: 402- 405).

115. There are some studies, however, that show that males and females can respond differently in general to the communications of a digital system. An example is also the study that was mentioned in chapter 4:

women are more susceptible to flattery from computers than men - they tend to respond negatively to it (Fogg & Nass, 1997).

116. The same can be said about the effects of the introduction of a new, more complex, risk scoring algorithm (the *PSA*) that was introduced in Kentucky in 2013. About this Stevenson (2018: 310) writes: "The switch from Kentucky's local risk assessment tool to the PSA did not result in any noticeable improvement in outcomes. There was a small increase in the use of non-financial bond, and essentially no effect on releases, FTAs, pretrial crime, or racial disparities in detention".

117. After the bill was implemented the racial gap - the difference between the proportion of black and white defendants - for bail without release jumped from 2 percentage point to 10 percentage point.

118. The particular study of Sanchez et al. (2014) focused on young adults' willingness operating agricultural vehicles to rely on the alarms of an automated collision avoidance system. They found that those with experience in operating agricultural vehicles or farming were more reluctant to rely on the system than those with little or no experience.

119. Hoffman describes one of the system's problems: "Of the first thirteen satellites launched in the test phase from 1972 to 1979, only seven worked for more than one hundred days. The satellites had to be launched constantly in order to keep enough of them aloft to monitor the American missile fields. They often just stopped sending data back to earth" (Hoffman, 2009: 9-10).

120. Also see Madhaven et al, 2006 c.f. Hoff & Bashir, 2015: 426.

121. Atoyan, Duquet and Robert (2006) demonstrated that trust in an intelligent data fusion system increased when the system was more transparent, clear, and provided more effective feedback. However, it is important to note that their findings were based on a study with only six participants, indicating that further research is necessary (c.f. Hoff & Bashir, 2015: 422-3).

122. In another variation, the participant was ordered to force a refusing victim's hand on the shock plate. The requirement of physical contact made that the level of obedience further decreased, to 30 percent (Milgram, 1974: 34).

123. In this context, see also the work of Kelman & Hamilton (1989: 336) on crimes of obedience.

124. Note that this is not necessarily so and that there are several studies on *automation bias*, as mentioned in this book's introductory chapter, highlighting that people tend to prefer computer output over other available sources of information (Mosier & Skitka, 1996; Skitka, Mosier & Burdick, 1999). But more insights are needed on what happens when digital technologies are actively challenged by a human authority figure.

125. The childcare benefits scandal involved the reliance by the Dutch tax authorities on a learning algorithm that used nationality, among other variables, to predict the risk of fraud. It served as a decision support system by flagging high-risk cases for further scrutiny. Tax employees manually checked the flagged applicants (Autoriteit Persoonsgegevens, 2020). The technology and tax officials turned out to be biased and disproportionately affected citizens with other nationalities (De Volkskrant, 2020). Many families were wrongly accused of benefits fraud based on their ethnic background and were required to repay large sums of money to the Dutch state, resulting in financial problems, bankruptcies, lasting mental health issues, divorces and broken homes (Geiger, 2021). In December 2020, the parliamentary report *Unprecedented Injustice* was published, leading to the resignation of the Dutch government several weeks later.

126. Picard calls these physical aspects of emotion *Sentic Modulation*. As she explains "sentic modulation, such as voice inflection, facial expression, and posture, is the physical means by which an emotional state is typically expressed, and is the primary means of communication human emotion" (Picard, 1997: 25).

127. Picard (1997:61-70) formulates five components of a system that has emotions:
    1. Emergent emotions and emotional behavior – displaying emotions.
    2. Fast Primary Emotions – having innate, quick and dirty emotions such as fear, surprise or anger.
    3. Cognitively Generated Emotions - having secondary (or social) emotions.
    4. Emotional Experience – the ability to recognise and label emotional behaviors, meaning:
        i. Cognitive awarness of emotions
        ii. Physiological awarness of emotions
        iii. Having subjective feelings
    5. Body-Mind Interactions – emotion's interaction with the mind and body.
128. This, of course, can all change. With developments in the field of Artificial Intelligence, it is impossible to predict their future emotional states and their ability to better understand and respond to us. When digital technologies become more emotionally intelligent, this will also surely impact their authoritative potential. This gives rise to all sorts of ethical dilemmas and discussions: is it a good thing or not to endow computers with emotional autonomy? However, most of these questions lead beyond this book – but see Picard (1997). In the next chapter there will be a brief section on the ethics of digital authority.
129. There is an increasing number of examples of digital toolkits and programs in Primary and Secondary schools. See for example the work of Bekker and colleagues (2015).

8. CONCLUSIONS

130. See paragraph 2.1.
131. Ian Kalman (2015) has argued that blaming a computer has also become a relevant aspect of the work of many frontline professionals. It can be taken as a modern effort at what Erving Goffman (1967) has called "face work". With this term Goffman refers to the various actions taken by an individual to express a positive "self-image" - or face - when interacting with others (1967: 5-23). He believes that in any interaction, individuals express a particular pattern – or a "line" – of the sort of person they are and their take on the other participants (1967:5). According to Kalman, maintaining a face - presenting a consistent image of oneself to others – is central to the work of good frontline-work and computer attributed referrals "offer a new means by which officers can articulate and save face in interactions" (2015: 8).
132. Nissenbaum's also identified "bugs" as a barrier to determining accountability. For her, the term covers a variety of software errors (i.e., modeling, design and coding errors) and considers them to be inevitable or "endemic to programming" (1994: 77). However, viewing bugs as "inevitable hazards of programming" not only highlights the vulnerability of many computing systems, it also makes holding people responsible problematic Nissenbaum says, since the harms and inconveniences that are caused cannot be helped - except in obvious cases of negligence (ibid). This implies that it would be unreasonable to actively hold developers, programmers, designers and other individuals responsible for the flaws in their systems.

    Nissenbaum's last barrier is "ownership without liability" (1994: 78): the idea that while companies aim to maximise their ownership over computer software, they also try to minimise responsibility for software-created problems. For this, their product is often accompanied by all sorts of disclaimers which make explicit that the producer is not liable to any form of damage that might follow from the use of their product. Nissenbaum mentions an Apple disclaimer as an example: "Apple makes no warranty or representation, either expressed or implied, with respect to software, its quality, performance, or merchantability, or fitness for a particular purpose. As a result, this software is sold 'as is,' and you, the purchaser are assuming the entire risk as to its quality and performance (…) In no event will Apple be liable for direct, indirect, special, incidental, or consequential damages resulting from any defect in the software or its documentation, even if advised of the possibility of such damages" (1994:78).
133. I asked the OpenAI model *ChatGPT* (2023) "What will be the future of Digital Authority". The quote is the summarising section of the answer that was provided on 21 November 2023.

SAMENVATTING

In de afgelopen decennia zijn we overspoeld met innovatieve, communicerende technologieën. Overal om ons heen vinden we digitale technologieën die informeren, adviseren, aanraden, beweren, instrueren en bevelen. Ze spelen een rol in allerlei omgevingen variërend van ons huis tot op de weg en van op werk tot in onze vrije tijd. In veel gevallen is het logisch om deze digitale communicaties te volgen. Ze zijn tenslotte ontworpen om uitdagingen aan te pakken en we hebben de waarde van computers al op veel gebieden gezien. Helaas is geen enkele technologie foutloos. Er kunnen altijd bugs, slechte algoritmen, biased data en invoerfouten zijn die ons op een dwaalspoor kunnen brengen. Maar hoewel de meesten van ons zich hier wel van bewust zijn, doen we geregeld maar al te graag wat ons digitaal wordt verteld.

Verhalen van mensen die in de problemen komen doordat zij automatisch een digitale technologie gehoorzamen komen veelvuldig in het nieuws – zo zijn er bijvoorbeeld talloze berichten over gevaarlijke situaties die ontstonden doordat mensen blindelings hun GPS-systeem volgden. Het vormt ook een populair thema voor satire. In de televisieshow *Little Britain* is er bijvoorbeeld een vrouwelijk karakter dat aan een bureau zit. Als haar klanten beschrijven wat ze willen, begint ze te typen met elke keer hetzelfde antwoord dat volgt: “computer zegt nee”. De uitdrukking is iconisch geworden voor het beschrijven van die situaties waarin mensen worden geleid door wat een computer hen ook maar vertelt.

Met de digitalisering van onze wereld zien we nu dat digitale systemen en omvangrijke databases betrokken raken bij steeds belangrijkere beslissingen. Verkeersovertredingen en fraude worden tegenwoordig op grote schaal digitaal geïdentificeerd en (semi-)automatisch afgehandeld. Bij het herkennen en voorkomen van allerhande risico’s zijn algoritmen en data centraal komen te staan. Zo zien we binnen de politie steeds meer voorbeelden van digitale systemen om criminaliteit te voorspellen en worden rechters wereldwijd steeds vaker voorzien van digitale risico-inschattingen bij het straffen en vrijlaten van verdachten. Daarnaast worden digitale technologieën ook gebruikt om kwetsbare personen die mogelijk in gevaar zijn beter te kunnen signaleren - zoals minderjarigen die het risico lopen om verwaarloosd of mishandeld te worden. De term *digital security governance* verwijst naar deze nieuwe benadering van werken aan veiligheid en het gebruik van allerhande digitale

technologieën voor het verzamelen, verwerken, opslaan, delen en toepassen van informatie. Het resultaat is dat digitale technologieën direct betrokken zijn geraakt bij beslissingen die een substantiële impact hebben op het welzijn van individuen en groepen – beslissingen over arrestaties, straffen, vrijlatingen, toegang en uitsluiting. Een ontwikkeling die op verschillende plekken geregeld leidt tot grote zorgen.

Tegen een achtergrond waarin organisaties voortdurend op zoek zijn naar geavanceerde technologieën voor veiligheid en andere wezenlijke kwesties, is de bereidheid van mensen om naar een computer te luisteren zowel intrigerend als verontrustend. Hoe ver kunnen we gaan? En waarom? Het volgen van digitale technologieën om vervolgens het verkeerde te doen is paradoxaal. Ten eerste omdat deze technologieën doorgaans zijn geïntroduceerd met het idee dat zij onze cognitie en besluiten juist verbeteren. Ten tweede omdat we "dingen" intuïtief nog vaak zien als neutrale en passieve instrumenten die we kunnen gebruiken om onze doelen te bereiken.

Eerder lieten wetenschappers al zien dat technologie veel meer is dan een instrument en ons op allerlei manieren kan beïnvloeden. Zo werd bijvoorbeeld duidelijk dat technologieën ons kunnen dwingen (Latour, 1992), overtuigen (Fogg, 2003), subtiel sturen (*nudging*) (Thaler & Sunstein, 2008) of misleiden (Castelfranchi, 2000). Toch bieden deze inzichten onvoldoende handvatten om bovengenoemde momenten van verregaande volgzaamheid te begrijpen. En ook de lens van menselijk vertrouwen in technologie lijkt een onvolledig beeld te geven van wat hier aan de hand is – zelfs als we de output van een computer niet geloven, negeren we soms toch ons eigen oordeel en doen we wat ons digitaal wordt gezegd. Het lijkt erop dat digitale technologieën hun rol als vertrouwde assistenten kunnen ontgroeien en ons op dominantere wijze kunnen sturen in wat we denken en doen. Dit brengt ons bij het idee van digitale autoriteit.

HET GEZAG DER DINGEN

Autoriteit (of *gezag*) is een omstreden begrip. Het wordt op veel verschillende manieren gebruikt en verwijst naar een diversiteit aan actoren – van experts en deskundigen tot degenen die formele ambten bekleden, zoals politieagenten en rechters. Wat zij gemeen hebben is dat hun communicaties mensen ertoe aan kunnen zetten te accepteren wat er ook tegen hen wordt gezegd. Autoriteiten beschikken over de capaciteit om dingen gedaan te krijgen. Het woord is afgeleid van het Latijnse *auctoritas*, dat in brede zin verwijst naar het vermogen om een dwingende kracht op mensen uit te oefenen "groter dan die van advies, maar minder dan die van bevel" (Mommsen, 1887 c.f. Arendt 1961: 123). Autoriteit is een speciaal soort invloed, verschillend van zowel dwang als overreding. Met autoriteit hoeft iemand niet te overtuigen, te dreigen of te dwingen om iets te bereiken. In plaats daarvan accepteren mensen wat autoriteiten zeggen uit vertrouwen en respect, of kunnen ze om strategische redenen doen alsof dit het geval is (Lincoln, 1994: 9). Autoriteit berust boven

alles op de vrijwillige erkenning van het subject dat naar de ander moet worden geluisterd. Het kan ook altijd worden betwist en ontkend. Maar zodra mensen autoriteit erkennen zullen zij geneigd zijn om alles dat tegen hen gezegd wordt te volgen, ongeacht hun eigen oordeel. Het lijkt dan net alsof we via een ander in de wereld denken en handelen.

In dit boek beschrijf ik autoriteit vanuit een interactionistisch perspectief, als iets dat voortkomt uit de uitwisselingen tussen ten minste twee actoren in een bepaalde setting. Iemand (of *iets*) heeft niet zomaar autoriteit. Het bestaat niet bij de gratie van individuele kwaliteiten zoals ambt, kennis, rijkdom, afkomst of charisma. Autoriteit is een gevolg, of een uitkomst van sociale interacties. Om dit idee verder te kunnen vatten onderscheid ik in hoofdstuk 3, een drietal conceptuele lagen. Ten eerste beschrijf ik autoriteit als een relationeel concept. Autoriteit impliceert een asymmetrische relatie waarin de een door de ander(en) niet als gelijke wordt beschouwd, maar in zekere zin als meerdere. Ten tweede is er een situationele laag. De erkenning van autoriteit ligt verankerd in de aard van de situatie. Er is een hele reeks signalen (items, gebaren, gedragingen, etc.) die autoriteit binnen een bepaalde setting overbrengt. Deze signalen kunnen mensen ertoe aanzetten een gehoorzame of accepterende rol aan te nemen. Ten derde bestaat er een culturele laag van autoriteit. Onze ideeën over autoriteit zijn cultureel gevormd. Gedurende ons leven, binnen het gezin, op school en later in de openbare en professionele ruimte, leren we wie we kunnen vertrouwen en dienen te respecteren. Op deze manier worden onze ideeën en verwachtingen van autoriteit geconditioneerd. Kortom, dit gelaagde perspectief op autoriteit benadrukt dat autoriteit wordt geconstrueerd in relaties, situaties en culturen. Hiermee verwordt autoriteit niet tot een statische of stabiele entiteit, maar staat de dynamiek en haar open essentie centraal. Dit doet recht aan het feit dat claims en erkenning van autoriteit in de loop van de tijd kunnen conflicteren en veranderen. En zo ontstaat ook ruimte voor reflectie op potentiële nieuwe autoriteiten.

Het concept van autoriteit is lange tijd perifeer gebleven binnen ons denken over technologie. Hoewel onze levens vol zitten met autoriteiten en de autoriteitsrelatie zo oud is als de menselijke beschaving zelf, is de term zelden gebruikt in relatie tot technologieën. In de hoofden van velen waren autoriteiten mensen zoals politieagenten, leraren, ouders of experts op een bepaald onderwerp – en niet dingen. Hoogstens werd gesproken over technologieën als instrument van menselijke autoriteiten of als factor die nieuwe autoriteitsrelaties tussen mensen tot stand kan brengen. Maar de autoriteit van technologieën zelf is echter vaak over het hoofd gezien of afgewezen. Toen nog nauwelijks sprake was van interactie tussen mensen en technologieën, was dit overigens heel logisch: er viel niets te gehoorzamen of accepteren.

EEN STUDIE NAAR DIGITALE AUTORITEIT

Nu digitale technologieën steeds meer kunnen, onderling verbonden zijn en toegang hebben tot een wereld aan data, zien we dat er nu wel af en toe gerichter gesproken wordt over de autoriteit van digitale technologieën. Ondanks dat dit vaak nog de vorm heeft van vluchtige verwijzingen, laat dit een voorzichtige verschuiving zien in het denken over de aard van autoriteit. Kennelijk is het niet meer zo vreemd om de term expliciet te koppelen aan technologische artefacten. Maar hoewel het toepassen van het idee van autoriteit op technologieën interessante wegen kan bieden voor het analyseren van hun sterke invloed op wat mensen denken en doen, brengt het ook nieuwe uitdagingen met zich mee en vereist het meer conceptuele helderheid. Want hoe moeten we deze digitale vorm van autoriteit begrijpen? Waar komt het vandaan en hoe is het opgebouwd? Hoewel technologieën nu misschien kunnen communiceren, ontberen ze nog steeds verstand en hebben ze geen intentie zoals mensen, laat staan enige vorm van vrijheid (Verbeek, 2011: 13).

De centrale vraag van dit proefschrift is:

In hoeverre kunnen digitale technologieën als autoriteiten worden beschouwd?

Het doel is om een bepaalde vorm van dwingende technologische invloed op menselijke beslissingen te conceptualiseren, om het fenomeen van mensen die voorbij hun eigen oordeel naar digitale technologieën luisteren beter te begrijpen. Een dergelijk nieuw begrip helpt ook om potentiële risico's zichtbaar te maken en deze te beperken. Hiervoor introduceer ik het idee van *digitale autoriteit*. Hiermee verwijs ik niet naar de aanwezigheid van menselijke autoriteit in de digitale wereld of gezaghebbende communicatie van mensen *via* computers. Ook gaat het hier niet over digitale technologieën die geprogrammeerd zijn om zelf autonome beslissingen te nemen in situaties waarin mensen afwezig zijn - denk bijvoorbeeld aan een computer die de werking van machines in een fabriek regelt of een volledig autonoom wapensysteem dat doelen aanvalt zonder enige menselijke tussenkomst. In plaats daarvan draait het bij digitale autoriteit om de technologieën zelf en de sterke invloed van hun communicaties op het menselijk handelen en denken. Specifieker verwijst het naar digitale technologieën en hun capaciteit om output te generen die door mensen als vanzelfsprekend wordt geaccepteerd uit vertrouwen, respect of hun bereidheid te doen alsof dit zo is. Zo kan digitale autoriteit zowel gaan over een bepaald soort invloed van digitale technologieën op mensen, als over de technologieën waarvan algemeen wordt aangenomen dat ze een dergelijke invloed kunnen uitoefenen. Dit omvat een breed scala aan digitale technologieën en richt niet uitsluitend op robots en geavanceerde vormen van artificiële intelligentie (AI). Daarnaast geldt dat digitale autoriteit niet noodzakelijkerwijs opzettelijk ontworpen hoeft te zijn. En hoewel ik in deze dissertatie primair verwijs naar het veiligheidsdomein vanwege de potentieel significante impact op de levens van mensen, kan digitale autoriteit overal voorkomen. Empirische onderzoeken en observaties vormen een

belangrijk fundament voor deze filosofische verkenning van digitale autoriteit en helpen om dit fenomeen verder te begrijpen.

Om de ontwikkeling en werking van digitale autoriteit te kunnen begrijpen ga ik in het tweede deel van het boek (hoofdstuk 4, 5 en 6) dieper in op de drie lagen van autoriteit: *relatie*, *situatie* en *cultuur*. Zij vormen de analytische kern van deze filosofische verkenning naar digitale autoriteit.

AUTORITEITSRELATIES MET TECHNOLOGIEËN

Hoofdstuk 4 richt zich op de relatie tussen mensen en technologieën. Centraal staat de vraag of hier ook autoriteitsrelaties zouden kunnen ontstaan, en hoe deze er dan uit zouden zien. Technologie laat zich niet goed begrijpen wanneer we deze benaderen als slechts een middel of instrument om onze doelen mee te bereiken. In plaats daarvan zette Don Ihde (1990) de relaties tussen mens en technologie centraal. Zo wordt zichtbaar hoe technologie mede vormgeeft aan onze ervaringen van de werkelijkheid en aan hoe wij hierbinnen aanwezig zijn. Ihde beschrijft diverse relaties tussen mens en technologie. Zo laat hij onder andere zien hoe we de wereld kunnen ervaren *via* een technologie - zoals bijvoorbeeld het geval is met een bril die door ons wordt "ingelijfd" en zo beïnvloedt hoe we alles om ons heen ervaren, of een thermometer die ons een bepaalde weergave van de wereld laat zien die wij vervolgens moeten interpreteren. Daarnaast beschrijft Ihde nog de zogeheten *alteriteitsrelatie* waarin technologische artefacten verworden tot quasi-anderen. Dit is het geval met veel digitale technologieën waar we nu mee *interacteren*.

Autoriteitsrelaties met technologieën gaan nog een stap verder. Hierbinnen worden menselijke percepties en handelingen sterk gevormd door een technologie die niet alleen een bron van interactie vormt, maar daarbij ook centraal staat in onze relatie met de werkelijkheid. Dit leert ons dat lang niet alle technologische artefacten potentiële autoriteiten zijn. We zullen een bril, een pen of een bijl niet als autoriteiten herkennen, omdat ze geen bron van interactie zijn. Maar zoals gezegd kunnen we met digitale technologieën een gevoel van "anderheid" ervaren, vanwege hun schijnbare autonomie en de interacties die ze veroorzaken. Meer specifiek zijn technologieën met hun digitalisering gaan communiceren. Lange tijd werd autoriteit gezien als een puur sociale relatie, juist omdat het alleen mensen waren die mededelingen deden die geloofd of gevolgd moesten worden. Maar met de komst van digitale technologie wordt de communicatie van dingen ook actief herkend door mensen – of het nu een pieptoon, een kunstmatige stem of een tekst op een scherm is. Digitale technologieën zijn een expliciete bron van communicatie en interactie geworden. Het bijzondere aan autoriteitsrelaties is echter dat zij fundamenteel asymmetrisch van aard zijn. Er zal dus een zekere mate van ongelijkheid tussen de mens en de technologie moeten worden ervaren die ons motiveert om te doen of te denken wat ons digitaal wordt verteld. In een autoriteitsrelatie is de invloed die technologieën uitoefenen

op menselijke percepties en gedrag zo dwingend dat deze kenmerkend is voor de relatie tussen mensen en hun wereld. We doen en denken dingen simpelweg omdat de technologie ons dat vertelt. Je kunt zeggen dat we in deze configuraties de wereld ervaren *door* of *volgens* een technologie.

Maar hoe zouden we het bestaan van deze asymmetrische relatie tussen mens en technologie kunnen begrijpen? Hoewel dit een vraag is die het hele boek beslaat, bespreek ik in hoofdstuk 4 enkele initiële elementen die kunnen worden beschouwd als constituerende onderdelen van digitale autoriteitsrelaties. Centraal staat hierbij het idee van *gedifferentieerde toegang* – autoriteit draait vaak om het idee dat iemand een bijzondere toegang heeft tot belangrijke kennis of tot specifieke machtsinstrumenten. Net als bij menselijke epistemische autoriteiten (zoals wetenschappers, experts en ervaringsdeskundigen) ervaren we bij digitale technologieën vaak dat ze toegang hebben tot een wereld aan unieke inzichten en informatie waar wij zelf niet bij kunnen. Ook kunnen we een bepaalde machtsongelijkheid ervaren in onze interacties met een digitale technologie – bijvoorbeeld zodra digitale systemen binnen een organisatie verplicht dienen te worden geraadpleegd bij het nemen van een beslissing of een technologie zo is geprogrammeerd dat het afwijkend gedrag simpelweg niet toelaat - ook wel *technoregulering* genoemd. Iemand kan het dan misschien niet eens zijn met de technologie, maar er uiteindelijk toch voor kiezen om te doen wat wordt opgedragen, omdat de technologie anders gehoorzaamheid wel zal afdwingen: de poort gaat niet open, het document wordt niet verzonden of het systeem reageert gewoon niet. Omdat technologieën geen emotionele wezens zijn die medelijden hebben met een ander, kan dit idee van mogelijke dwang extra krachtig zijn: technologieën maken namelijk geen uitzonderingen.

Deze verschillen in toegang tot kennis en formele macht maken dat ook binnen digitale autoriteitsrelaties overtuiging en dwang op de achtergrond een belangrijke rol spelen. De latente aanwezigheid van deze andere manieren van invloed geeft de asymmetrie van autoriteitsrelaties tussen mensen en technologieën vorm en zorgt voor acceptatie, vertrouwen, respect of de bereidheid te doen alsof dit zo zou zijn. Het leidt tot de veronderstelling dat in gevallen van twijfel of onwil de technologie in staat is om ons te overtuigen door middel van de juiste argumenten of simpelweg kan dwingen om mee te werken. In de praktijk zijn technologieën natuurlijk vaak niet geprogrammeerd om zichzelf uit te leggen en zijn ze nog vaak afhankelijk van anderen om overtuigende argumenten aan te dragen voor hun output of om dwang uit te oefenen in tijden van verzet. Maar waar het om draait is het idee dat mensen denken dat er logischerwijs bewijs bestaat voor digitale claims of dat de technologie in staat is om medewerking eventueel af te dwingen. Het is deze veronderstelling die mensen vooraf overtuigt dat naar de computer geluisterd dient te worden. Echter, als belangrijke herinnering, wanneer dwang of overreding daadwerkelijk wordt toegepast, zou er geen sprake zijn van digitale autoriteit maar slechts van digitale dwang of overreding.

DIGITALE AUTORITEITSSITUATIES

Hoofdstuk 5 gaat vervolgens in op mogelijke autoriteitssituaties met digitale technologieën. Hannah Arendt (1963) en Stanley Milgram (1974) betoogden decennia geleden al dat gezag berust op de aard van de situatie. In het kort: als mensen een autoriteitssituatie herkennen zullen zij een gehoorzame rol aannemen en doen of denken wat hen wordt gezegd. Om dit beter te begrijpen beschreef ik de noties van definitie van de situatie en scripts. Wanneer mensen in een nieuwe situatie terechtkomen, vragen ze zich af: wat is er aan de hand? En hoe moet ik me hier gedragen? (Goffman, 1986; Van den Berg, 2009). Om deze vragen te beantwoorden, vertrouwen mensen veelal op aangeleerde scripts. Scripts zijn kennispakketten waarmee mensen door hun dagelijks leven kunnen navigeren. Ze specificeren de rollen van personen en objecten binnen een bepaalde situatie en bevatten bepaalde actiepatronen (Schank & Abelson, 1977). Een omgeving met de juiste signalen zal vervolgens een specifiek script activeren en bepaalde ervaringen en gedragingen stimuleren. Zo kunnen we ook begrijpen hoe mensen autoriteit herkennen en hierop reageren: met de juiste aanwijzingen kunnen scripts worden geactiveerd die mensen ertoe aanzetten iets als een autoriteitssituatie te definiëren om vervolgens een accepterende, eerbiedige of zelfs onderdanige rol aan te nemen. Er bestaan tal van gezaghebbende items, plaatsen, symbolen of gedragingen die als situationele signalen fungeren die individuen helpen autoriteit te herkennen en daarmee een mentale shortcut bieden om gepast te reageren. Zo wordt duidelijk dat degenen die autoriteit claimen en zich op een bepaalde manier aan ons presenteren, een situatie effectief trachten te vormen voor de ander: met behulp van de juiste signalen stimuleren ze iemand om de situatie te definiëren als autoriteitssituatie.

Maar wat zegt dit over digitale autoriteit? Vooralsnog ging het immers enkel over autoriteitssituaties met andere mensen. Toch lijken ook digitale technologieën gebruik te maken van dezelfde sociale scripts en op vergelijkbare wijze autoriteitssituaties voor ons te kunnen creëren. Uit onderzoek komt naar voren dat digitale technologieën sterke sociale reacties bij ons kunnen oproepen – het soort reacties dat we traditioneel reserveren voor interacties met andere mensen (Reeves & Nass, 1996; Turkle, 2005). En hiervoor hoeven de sociale signalen van technologieën niet bijzonder geavanceerd te zijn en hoeven mensen zich niet bewust te zijn van hun sociale gedragingen. Het is daarom misschien niet langer nodig om een sterk onderscheid te maken tussen menselijke en technologische autoriteitssituaties: door de juiste presentatie kunnen technologieën op een zeer vergelijkbare wijze centraal komen te staan in autoriteitssituaties als menselijke autoriteiten dat doen. Zo zien we nu dat digitale technologieën traditionele autoriteitsrollen vervullen die voorheen werden vervuld door menselijke gezagsdragers, zoals artsen, leraren, deskundigen of politiefunctionarissen. Ook worden er allerlei titels gegeven die dit nog eens benadrukken: technologieën die bijvoorbeeld als “expert” of “specialist” worden bestempeld. En soms zijn er meer verfijnde signalen, denk aan een bepaalde taalstijl of autoriteitssymbolen zoals certificaten of logo’s.

Al deze signalen die een rol spelen bij de constructie van menselijke autoriteitssituaties kunnen in interacties met digitale technologieën vergelijkbare reacties oproepen.

Kortom, digitale technologieën hoeven er niet precies hetzelfde uit te zien en te handelen als mensen, om subjecten ertoe te brengen autoriteitssituaties te herkennen en er dienovereenkomstig op te reageren. En het maakt niet uit of hun kennis of bevoegdheden echt zijn – of technologieën nu legitieme autoriteiten zijn of niet. Waar het om gaat is de gezaghebbende indruk die ze kunnen maken die een autoriteitsscript kan triggeren. In deze situaties, een belangrijke toevoeging, functioneren digitale technologieën niet als proxy voor menselijke autoriteit. Uit verschillende experimenten is gebleken dat individuen zich niet richten op degenen achter de computer - de programmeurs of hun bazen. Integendeel, teruggrijpend op de eerdere discussie over de relaties tussen mens en technologie: het is de digitale technologie waarop we onze sociale reacties richten.

EEN CULTUUR VOOR DIGITALE AUTORITEIT

Hoofdstuk 6 richt zich op de vraag in hoeverre er plaats is voor digitale autoriteit in hedendaagse culturen. Er bestaan aanzienlijke verschillen tussen culturen in de mate waarin mensen hiërarchieën en ongelijkheid in sociale relaties leren accepteren (Hofstede et al, 2010). Sommige culturen zijn meer op autoriteit georiënteerd dan anderen. Ook groeien mensen op in verschillende omgevingen met eigen ideeën over technologische innovatie. Toch zijn de voorbeelden van mensen die blind hun digitale technologieën volgen wereldwijd verspreid. Als we vervolgens inzoomen op hoe *binnen* samenlevingen ideeën van digitale autoriteit zouden kunnen postvatten ontdekken we twee conflicterende technologische narratieven. Enerzijds leren mensen over hoe technologie elk probleem kan oplossen en alles makkelijker kan maken. Anderzijds is er doorlopend aandacht voor de gevaren en risico's van technologische innovatie. Binnen het veiligheidsdomein zien we bijvoorbeeld hoe digitale technologieën vaak worden beschouwd als waardevolle oplossing voor misdaad en allerhande dreigingen, maar ook als een potentieel gevaar voor het individu. Onze moderne culturen zitten vol met elementen van deze twee fundamenteel verschillende verhalen van acceptatie en afwijzing, van moderne technologie als iets goeds of slechts. Wat zij gemeen hebben is een deterministische kijk op de invloed van nieuwe technologieën op samenlevingen – er wordt aangenomen dat zij de wereld kunnen redden of ruïneren. Het benadrukken van de vormende of transformerende capaciteiten van technologieën kan zorgen voor ontzag en de constructie van digitale autoriteitsrelaties stimuleren. Deze positieve en negatieve ideeën zullen er echter ook toe leiden dat mensen op tegenstrijdige manieren reageren op digitale technologieën. Over het algemeen is de moderne samenleving dus een omgeving van ambivalentie die digitale autoriteit zowel kan stimuleren als beperken.

Maar er is nog een diepere culturele laag. Er is al veel geschreven over dat de moderne tijd wordt gekenmerkt door het toegenomen bewustzijn van mensen over de onzekerheid en complexiteit van hun wereld. Dit heeft onder andere te maken met de erosie van traditie en religie als ook met het feit dat allerhande nieuwe gevaren steeds meer het gevolg blijken van menselijk handelen. Mensen maken zich hierdoor meer zorgen en voelen een groeiende verantwoordelijkheid voor wat de toekomst brengt. Aanvankelijk gaf de imposante ontwikkeling van wetenschap en technologie in moderne samenlevingen mensen de indruk van meer zekerheid en controle. Maar al snel leerde men dat binnen de wetenschap vaak onenigheid bestaat en kennis voortdurend wordt herzien. Ook dienden technologische catastrofes zoals kernrampen en grootschalige vervuiling zich aan. Zo creëerde de groeiende maatschappelijke rol van wetenschap en technologie uiteindelijk vooral een reeks nieuwe onzekerheden en angsten. Ulrich Beck (1992a) en Anthony Giddens (1991) stelden dat zo een risicocultuur of -maatschappij ontwikkelde, waarin een algemeen klimaat van onzekerheid heerst en een obsessie met toekomstige schade die ons zou kunnen overkomen.

Onder deze moderne omstandigheden zijn mensen steeds meer bezig met zoeken naar en creëren van zekerheid. Individuen nemen een *reflexieve* houding aan bij het maken van keuzes en handelen in hun wereld. Een onderdeel hiervan is dat men voortdurend actief is met het monitoren wie er kan worden vertrouwd en gevolgd. In een samenleving na religie en traditie, waar risico's nu worden ervaren als menselijke verantwoordelijkheden, moeten individuen op zoek naar nieuwe relevante expertise en betrouwbare anderen om naar te luisteren. Zo heeft het bewustzijn van de beperkingen van wetenschappelijke kennis en twijfel aan traditioneel gezag ruimte gemaakt voor de opkomst van andere autoriteiten. Individuen staan in de moderne tijd misschien wel meer dan ooit open voor nieuwe autoriteiten die hen de wereld kunnen helpen leren kennen en controleren - die ongemakkelijke gevoelens van onzekerheid en twijfel kunnen helpen tegengaan. Voor de meesten zal dit betekenen dat zij uiteindelijk hun weg zullen vinden door te vertrouwen op een reeks verschillende autoriteiten. En zodra digitale technologieën worden gepresenteerd als effectieve oplossingen tegen onzekerheid en complexiteit verworden moderne samenlevingen dan ook tot een vruchtbare omgeving voor het ontstaan van digitale autoriteit.

GRENZEN

Digitale autoriteit is echter geen gegeven. In het laatste deel van deze dissertatie beschrijf ik dat er veel zaken zijn die de constructie van digitale autoriteit kunnen belemmeren. Zo bekleden deze technologieën bijvoorbeeld geen officiële ambten en hebben zij niet dezelfde formele bevoegdheden die de positie van veel menselijke autoriteiten versterken. Daarnaast zijn er ook nog steeds tal van gebieden waar individuen over voldoende kennis

en ervaring beschikken om op hun eigen oordeel te vertrouwen. En helpen fouten van een technologie ook niet bij de ontwikkeling van digitale autoriteit. Digitale autoriteit is ook niet vrijgesteld van het reflexieve proces waarvan zij zelf ook het product is - mensen kunnen altijd op zoek gaan naar nieuwe bronnen van zekerheid.

Daarnaast geldt dat digitale autoriteit wordt begrensd doordat bij technologieën vooralsnog het vermogen ontbreekt om momenten van potentiële crisis waar te nemen en hier effectief op te reageren. Dit zijn de momenten waarop mensen, expliciet of impliciet, twijfelen of nee zeggen tegen een technologie. Technologieën ontberen een geest en hebben geen intentie of autonomie zoals mensen dat hebben, waardoor ze slechts een beperkt actierepertoire hebben om met weerstand en uitdaging om te gaan. Vooralsnog reageren de meeste technologieën slechts op de manier waarop ze zijn geprogrammeerd, waardoor ze vaak alleen maar in staat zijn hun output te herhalen aan een onwillig menselijk subject. Digitale technologieën zijn over het algemeen afhankelijk van anderen om een publiek weer voor zich te winnen – bijvoorbeeld van mensen die potentiële voordelen van naleving kunnen aangeven of door de negatieve gevolgen van weigering onder de aandacht te brengen. Dit maakt digitale autoriteit over het algemeen kwetsbaar en beperkt. Want wat als die anderen afwezig zijn?

Nu digitale innovatie steeds meer samenvalt met de voortdurende ontwikkeling van artificiële intelligentie, kunnen de grenzen van digitale autoriteit de komende decennia makkelijk verschuiven. Digitale technologieën zullen zich steeds gezaghebbender presenteren en steeds veerkrachtiger worden bij menselijk verzet. Tegelijkertijd laten de eerste digitale autoriteiten zich nu al duidelijk zien. Hoewel digitale technologieën ons de mogelijkheid bieden om gerichte beslissingen te nemen waar we anders niet toe in staat waren geweest, hebben ze ook tot nieuwe incidenten en risico's geleid. Een rode draad in deze dissertatie zijn de verhalen waarin mensen duidelijk te ver gaan in het volgen van digitale autoriteiten. Door individuele gevoelens van doelmatigheid (*efficacy*) en bevoegdheid (*empowerment*) te vergroten kunnen sommige potentiële valkuilen van digitale autoriteit worden verzacht. Zo kunnen bijvoorbeeld onderwijs, training en het uitdragen van duidelijke richtlijnen individuele oordeel- en besluitvorming stimuleren en mogelijk maken. Ten slotte blijft het belangrijk om te erkennen dat er altijd dingen kunnen gebeuren in omgevingen waarin er soms eenvoudigweg geen rationeel alternatief is voor het volgen van een digitale technologie. Individuen, organisaties en samenlevingen moeten zich hierop inrichten en structureel aandacht hebben voor de eventuele schadelijke impact van digitale autoriteit.

CONCLUSIE

In het laatste hoofdstuk concludeer ik dat digitale technologieën ons dwingen om veel opvattingen over de aard van artefacten en mensen te heroverwegen. Mensen zijn niet uitsluitend “gebruikers” en dingen niet slechts instrumenten. Dat blijkt ook uit die gevallen waarin we technologie respecteren en volgen. Het concept autoriteit helpt om de verregaande bereidheid van mensen om te luisteren naar een digitale technologie beter te begrijpen. En een analyse van de verschillende autoriteitslagen kan het gezaghebbende potentieel van specifieke technologieën verder verduidelijken. Digitale autoriteit is dan het resultaat van het samenvallen van een digitale autoriteitsrelatie, situatie en cultureel specifieke geconditioneerde verwachtingen en ideeën. Dit zorgt ervoor dat interactie met een digitale technologie zorgt voor vertrouwen, respect, acceptatie of gehoorzaamheid onder mensen.

Autoriteit is dus niet dood, zoals we soms lezen, maar lijkt zich te verplaatsen naar nieuwe plekken, zoals de wereld van de technologie. Dit biedt kansen voor betere besluiten maar zorgt dus ook voor wezenlijke nieuwe gevaren. Zeker op plaatsen waar de uitwisseling tussen mens en technologie de basis vormt voor belangrijke beslissingen die het welzijn van het individu direct raken - denk aan het veiligheidsdomein – is voorzichtigheid geboden. We zullen daarom na moeten blijven denken over verantwoorde ontwikkeling van digitale autoriteit in onze wereld. De laatste paragrafen zijn hier een eerste aanzet toe. Zo bespreek ik de waarde van verder empirisch onderzoek en het belang van de ontwikkeling van een ethiek van digitale autoriteit. Ook sta ik stil bij de complexiteit van verantwoordelijkheid, aansprakelijkheid en schuld in een context van digitale autoriteit.

L LITERATURE

Abelson, R. P. (1981),'Psychological status of the script concept'. *American Psychologist,* 36(7), pp.715–729.

Achterhuis, H. (1997), *Van stoommachine tot cyborg; denken over techniek in de nieuwe wereld*. Amsterdam*:* Ambo.

Adam, B. (1996), 'Detraditionalization and the Certainty of Uncertain Futures', in Heelas, P., Lash, S. & Morris, P. (eds), *Detraditionalization. Critical Reflections on Authority and Identity* (pp. 134-48). Oxford: Blackwell.

Adorno, T., Frenkel-Brunswick, E., Levinson, D. & Sanford, N., (1950). *The Authoritarian Personality*. New York: Harper.

AI Impacts (2022), *2022 Expert Survey on Progress in AI*. Online via https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/ (Accessed January 2024).

Akrich, M. (1992),'The De-Scription of Technical Objects', in Bijker, W. and Law, J. (eds.), *Shaping Technology/ Building Society: Studies in Socio-technical change* (pp. 205-224). Cambridge: MIT Press.

Albarracin, M., Constant, A., Friston, K. & Ramstead, M. (2021),'A variational approach to scripts'. *Frontiers in Psychology*, 12(1), pp.1-15.

Albrechtslund, A. (2007),'Ethics and technology design'. *Ethics and Information Technology*, 9(1), pp. 63-72.

Albright, A. (2019),'If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions'. *The John M. Olin Center for Law, Economics, and Business Fellows' Discussion Paper Series*, *85*.

Algemeen Dagblad (2024), Gevangene die 401 dagen te vroeg werd vrijgelaten is nu spoorloos. Online via https://www.ad.nl/binnenland/gevangene-die-401-dagen-te-vroeg-werd-vrijgelaten-is-nu-spoorloos~aaea6689/?referrer=https%3A%2F%2Fwww.google.com%2F (Accessed March 2024).

Alon-Barkat, S., & Busuioc, M. (2023),'Human–AI interactions in public sector decision making:"automation bias" and "selective adherence" to algorithmic advice'. *Journal of Public Administration Research and Theory*, 33(1), pp.153-169.

Altemeyer, R.A. (1981), *Right-wing Authoritarianism*. Winnipeg: University of Manitoba Press.

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.

Amnesty International, (2019), *We Sense Trouble: Automated discrimination and mass surveillance in predictive policing in the Netherlands*. Online via https://www.amnesty.org/en/documents/eur35/2971/2020/en/ (Accessed January 2024).

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016),'Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks'. *ProPublica*. Online via https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Accessed January 2024).

Anttiroiki, A., Bailey, S. & Valkama, P. (2012),'Innovations in public governance in the western world', in Anttiroiki, A., Bailey, S & Valkama, P. (eds.), *Innovation and the public sector* (pp. 1-22). Amsterdam: IOS Press.

Arendt, H. (1961), *Between past and future: Eight exercises in Political Thought*. New York: Viking Press.

Arendt, H. (1963), *Eichmann in Jerusalem: a Report on the Banality of Evil*. New York: Viking Press.

Arendt, H. (1970), *On Violence*. New York: Hartcourt Brace Jovanovich Publishers.

Asworth, A. & Zedner, L. (2014), *Preventive Justice*. Oxford: Oxford University Press.

Austin, V.L. (2010), *Up with Authority: Why we need authority to flourish as human beings*. London: T&T Clark.

AT5 (2018),'Apple-Navigate stuurt fietsende toeristen door de Piet Heintunnel'. Online via https://www.at5.nl/artikelen/184990/apple-navigatie-stuurt-fietsende-toeristen-door-de-piet-heintunnel (Accessed January 2024).

Atoyan H., Duquet J. & Robert J. (2006),'Trust in new decision aid systems'. *Proceedings of the 18th International on l'Interaction Homme-Machine*, pp. 115–122.

Autoriteit Persoonsgegevens (2020), *Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag*. Online via https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf (Accessed January 2024).

Axell, C. (2018),'Technology and children's literature', in De Vries, M. (ed.), *Handbook of Technology Education* (pp. 895-911). Cham: Springer.

Axell, C. & Boström, J. (2021),'Technology in children's picture books as an agent for reinforcing or challenging traditional gender stereotypes'. *International journal of technology and design education*, 31(1), pp. 27-39.

Bandura, A. (1978),'Self-efficacy: toward a unifying theory of behavioral change'. *Psychological review*, 84(2), pp. 139-161.

Bauman, Z. (2012), *Liquid modernity*. Cambridge: Polity press.

Bartneck, C. & Hu, J. (2008),'Exploring the abuse of robots'. *Interaction Studies*, *9*(3), pp. 415-433.

Barton, T. (2011), *Defying Defcon One*. RT News. Online via https://www.youtube.com/watch?v=2pJ4Eu5VFOA&t=5s (Accessed January 2024).

BBC (2013a), *Apple Maps flaw results in drivers crossing airport runway*. Online via https://www.bbc.com/news/technology-24246646 (Accessed January 2024).

BBC (2013b*), Stanislav Petrov: The man who may have saved the world*. Online via https://www.bbc.com/news/world-europe-24280831 (Accessed January 2024).

BBC (2020), *Hawk-Eye apologizes after goalline technology error denies Sheffield United*. Online via https://www.bbc.com/sport/football/53086360 (Accessed January 2024).

BBC (2024), Airline held liable for its chatbot giving passengers bad advice – what this means for travelers. Online via https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know (Accessed February 2024).

Beck, U. (1992a). *Risk society: Towards a new modernity*. London: Sage.

Beck, U. (1992b). From industrial society to the risk society: Questions of survival, social structure, and ecological enlightenment. *Theory, culture & society*, *9*(1), pp.97-123.

Beck, U. (1998),'Politics of Risk Society', in Franklin, J. (ed.), *The Politics of Risk Society* (pp.9-23). Cambridge: Polity Press.

Beck, U. (2003), *An interview with Ulrich Beck on fear and risk society*. by Yates, J. in *The Hedgehog Review*, 5(3), pp. 96-108.

Bedner, A. W. (2002). Judicial corruption: Some consequences, causes and remedies. *CILC Seminar in Tribute to Jan van Olden*, *Center for International Legal Cooperation*. Online via https://scholarlypublications.universiteitleiden.nl/access/item%3A2931169/download (Accessed January 2024).

Bekker, T., Bakker, S., Douma, I., Van Der Poel, J. & Scheltenaar, K. (2015),'Teaching children digital literacy through design-based learning with digital toolkits in schools'. *International Journal of Child-Computer Interaction*, 5 (1), pp. 29-38.

Bell, D. (1970),'Authority'. *Royal Institute of Philosophy Lectures*, *4*, pp. 190-203.

Bernstein, P. L. (1996). *Against the gods: The remarkable story of risk*. New York: Wiley.

Berdichewsky, D. & Neuenschwander, E. (1999),'Towards an ethics of persuasive technology'. *Communications of the ACM,* 42(5), pp.51-58.

Bevir, M. (2009), *Key Concepts in Governance*. London: Sage.

Bick, M., & Kummer, T. F. (2008),'Ambient intelligence and Ubiquitous computing', in Adelsberger, H., Pawlowski, J. & Sampson, D. (eds.), *Handbook on Information Technologies for Education and Training* (pp. 79-100). Berlin: Springer.

Bickman, L. (1974), 'The Social Power of a Uniform'. *Journal of Applied Social Psychology*, 4 (1), pp. 47-61.

Bidle, J. (2018), *New Dark Age: Technology and the end of the Future*. London: Verso Books.

Bingham, C. (2008), *Authority is relational: Rethinking educational empowerment*. New York: SUNY Press.

Biros, D. P., Fields, G. & Gunsch, G. (2003),'The effect of external safeguards on human-information system trust in an information warfare environment'. *International Conference on System Sciences,* pp.1-10.

Blass, T. (1991),'Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions'. *Journal of Personality and Social Psychology*, 60(3), pp. 398-413.

Blass, T. (1995),'Right-wing authoritarianism and role as predictors of attributions about obedience to authority'. *Personality and Individual differences*, 19(1), pp. 99-100.

Blass, T. (1996),'Attribution of Responsibility and Trust in the Milgram Obedience Experiment 1'. *Journal of Applied Social Psychology*, 26(17), pp.1529-1535.

Blass, T. (1999),'The Milgram Paradigm After 35 Years: Some Things We Now Know About Obedience to Authority'. *Journal of Applied Social Psychology*, 29(5), pp.955-978.

Bocchiaro, P., & Zimbardo, P. (2008),'Deciding to defy unjust authority: An experimental investigation'. In Zimbardo, P. (ed.), *Acta Zimbardo Foundation*, (pp. 29-56). Cammarata, Italy.

Bock, D. C., & Warren, N. (1972),'Religious Belief as a Factor in Obedience to Destructive Commands'. *Review of Religious Research*, 13(1), pp.185-191.

Bosworth, B. (1999),'Augustus, the Res Gestae and Hellenistic Theories of Apotheosis'. *The Journal of Roman Studies*, 89 (1), pp. 1-18.

Bovens, M. & Zouridis, S. (2002),'From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control'. *Public administration review*, *62*(2), pp.174-184.

Bostrom, N. (2002),'Existential risks: Analyzing human extinction scenarios and related hazards'. *Journal of Evolution and Technology*, 9(1), pp.1-30.

Bostrom, N. (2006), *AI set to exceed human brain power*. CNN. Online via http://edition.cnn.com/2006/TECH/science/07/24/ai.bostrom/ (Accessed January 2024).

Bostrom, N. (2013),' Existential risk prevention as global priority'. *Global Policy*, 4(1), pp.15-31.

Bostrom, N. & Sandberg, A. (2009),'Cognitive enhancement: methods, ethics, regulatory Challenges'. *Science and Engineering Ethics*, 15(3), 311-341.

Botsman, R. (2017). *Who Can You Trust?: How Technology Brought Us Together–and Why It Could Drive Us Apart*. London: Penguin.

Bourdieu, P. (1990a), *In Other Words: Essays towards a reflexive sociology*. Stanford: Stanford University Press.

Bourdieu, P. (1990b), *The Logic of Practice*. Translated by Nice, R., Stanford: Stanford University Press.

Bozinoff, L. (1982),'A Script Theoretic Approach to Information Processing: an Energy Conservation Application'. Advances in Consumer Research, 9(1), pp.481-486.

Bracken, P. (2012), *The second nuclear age: Strategy, danger, and the new power politics*. New York: St. Martin's Press.

Breakwell, G.M (2014), *The psychology of risk (2nd edition).* Cambridge: Cambridge University Press.

Brey, P. (2005),'Freedom and privacy in ambient intelligence'. *Ethics and Information Technology*, 7(3), pp.157-166.

Broeders, D., Schrijvers, E., Van Braker, R.E., de Hoog, J., Van der Sloot, B. & Hirsch-Ballin, E. (2016), *Big Data in een vrije en veilige samenleving*. Wetenschappelijke Raad voor Regeringsbeleid. Amsterdam: Amsterdam University Press.

Broeders, D., Schrijvers, E., Van der Sloot, B., Van Brakel, R., de Hoog, J. & Hirsch-Ballin, E. (2017),'Big Data and Security Policies: Towards a framework for regulating the phases of analytics and use of Big Data'. *Computer Law & Security Review*, 33(3), pp.309-323.

Broussard, M. (2018), *Artificial Unintelligence: How computers misunderstand the World*. Cambridge: MIT Press.

Brownsword, R. (2004),'What the world needs now: Techno-regulation, human rights and human dignity'. in Bronsword, R. (ed.), *Global Governance and the Quest for Justice Volume 4* (pp.203-234). Oxford: Hart Publishing.

Brownsword, R., & Yeung, K. (eds.). (2008). *Regulating technologies: legal futures, regulatory frames and technological fixes*. London: Bloomsbury Publishing.

Buffat, A. (2015),'Street-level bureaucracy and e-government'. *Public Management Review*, 17(1), pp.149-161.

Burley, P. M. & McGuinness, J. (1977),'Effects of social intelligence on the Milgram paradigm'. *Psychological Reports*, 40(3), pp. 767-770.

Burger, J. M. (2009),'Replicating Milgram: Would people still obey today?'. *American Psychologist*, 64(1), pp.1-11.

Busch, P. A., & Henriksen, H. Z. (2018),'Digital discretion: A systematic literature review of ICT and street-level discretion'. *Information Polity*, 23(1), pp.3-28.

Carlson, M. (2017), *Journalistic authority: Legitimating news in the digital era*. New York: Columbia University Press.

Byrne, J. & Marx, G. (2011),'Technological innovations in crime prevention and policing. A review of the research on implementation and impact'. *Journal of Police Studies*, 20(3), pp. 17-40.

Caporael, L. R. (1986),'Anthropomorphism and mechanomorphism: Two faces of the human machine'. *Computers in Human Behavior*, 2(3), pp. 215-234.

Cardon, P. & Marshall, B. (2008),'National Culture and Technology Acceptance: The impact of uncertainty avoidance'. *Issues in Information Systems*, 9(2), pp.103-110.

Carter, A. (1979). *Authority and Democracy.* London: Routledge.

Casero-Ripollés, A. (2021),' Influencers in the political conversation on twitter: Identifying digital authority with big data'. *Sustainability*, *13*(5), 2851.

Castelfranchi, C. (2000),'Artificial liars: Why computers will (necessarily) deceive us and each other'. *Ethics and Information Technology*, 2(2), pp.113-119.

Center for AI Safety (2023), *Statement on AI Risk: AI experts and public figures express their concern about AI risk*. Online via https://www.safe.ai/statement-on-ai-risk (Accessed January 2024).

Chanenson, S. & Hyatt, J. (2016), *The use of risk assessment at sentencing: Implications for research and policy*. Villanova University Charles Widger School of Law. Online via https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2961288 (Accessed January 2024).

Cialdini, R. B., (2007), *Influence: The Psychology of Persuasion*. New York: Collins.

Citron, D. K. (2008),'Open Code Governance'. The *University Chicago Legal Forum*, pp.355-387.

Clark, A. (1999),'An embodied cognitive Science?. *Trends in cognitive sciences*, 3(9), pp.345-351.

CNN (2008), *Airline captain, lawyer, child on terror 'watch list'*. Online via https://edition.cnn.com/2008/US/08/19/tsa.watch.list/index.html (Accessed January 2024).

CNN (2012), No-Fly 'glitch' has toddler removed from place. Online via https://edition.cnn.com/2012/05/10/travel/no-fly-toddler/index.html#:~:text="TSA%20was%20called%20to%20the,flight%2C%20saying%20they%20felt%20humiliated (Accessed January 2024).

CNN (2019), *Nearly 100 drivers followed a Google Maps detour – and ended up in an empty field*. Online via https://edition.cnn.com/2019/06/26/us/google-maps-detour-colorado-trnd/index.html (Accessed January 2024).

Cohen, J.L. (1982), *An Essay on Belief and Acceptance*. Oxford: Oxford University Press.

Coiera, E. (2007),'Putting the technical back into socio-technical systems research'. *International journal of medical informatics*, 76(1), pp.98-103.

Collins, H., & Evans, R. (2008),'You cannot be serious! Public understanding of technology with special reference to "Hawk-Eye"'. *Public Understanding of Science*, 17(3), pp.283-308.

Cooper, M. D. (2000). Towards a model of safety culture. *Safety science*, 36(2), pp. 111-136.

Cormier, D., Newman, G., Nakane, M., Young, J. E. & Durocher, S. (2013).,'Would you do as a robot commands? an obedience study for human-robot interaction. *International Conference on Human-Agent Interaction*, pp. 1-8.

Cortada, J. (2008), *The Digital Hand: How computers changed the work of American public sector industries*. New York: Oxford University Press.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Cukier, K. (2010), *Data, data everywhere*. The Economist. Online via https://www.economist.com/special-report/2010/02/27/data-data-everywhere (Accessed January 2024).

Cukier, K. & Mayer-Schonberger, V. (2013), *Big Data: A revolution that will transform who we live, work, and think*. Boston: Houghton Mifflin Harcourt.

Danaher, J. (2016),'The Threat of Algocracy: Reality, resistance and accommodation'. *Philosophy & Technology*, 29(3), pp. 245-268.

Daugherty, W., Levi, B., & Vons Hippel, F. (1986),'The consequences of" limited" nuclear attacks on the United States'. *International Security*, 10(4), pp.3-45.

Davies, D. (2024), *The Unaccountability Machine: Why Big Systems make Terrible Decisions - and How the World Lost Its Mind*. London: Profile Books.

Davis, G. A., Paullet, K. L., Houck, M. & Swan, T. (2010),'Does the technology portrayed in television crime shows have an effect on potential jurors?'. *Issues in Information Systems*, 11(1), pp.154-163

De George, R. T. (1985). *The nature and limits of Authority*. Lawrence: University Press of Kansas.

De Jouvenel, B. (1957). *Sovereignty: An inquiry into the political good*. Chicago: University of Chicago Press.

De La Mettrie, J.O. (1747), *Man a Machine*. Translated by Bussey (2016). G.C. & Calkins, M.W. Rockville: Wildside Press.

De Mul, J. & Van den Berg, B. (2011),'Remote control: Human autonomy in the age of computer-mediated agency', in Hildebrandt, M. & Rouvroy, A. (eds.), *Law, Human Agency and Autonomic Computing* (pp. 62-79). London: Routledge.

De Mul, J. (2013),'The (Bio)Technological Sublime in Aesthetics in action'. Diogenes, 59(1), pp.32:40.

De Standaard (2013), *Oma volgt gps naar Brussel en stopt pas in Kroatië*. Online via https://www.standaard.be/cnt/dmf20130112_00431308 (Accessed January 2024).

De Volkskrant (2020), *Belastingdienst schuldig aan structurele discriminatie van mensen die toeslagen ontvingen*. Online via https://www.volkskrant.nl/nieuws-achtergrond/belastingdienst-schuldig-aan-structurele-discriminatie-van-mensen-die-toeslagen-ontvingen~baebefdb/ (Accessed January 2024).

De Visser, E. J., Krueger, F., McKnight, P., Scheid, S., Smith, M., Chalk, S. & Parasuraman, R. (2012),'The world is not enough: Trust in cognitive agents'. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 56(1), pp. 263–267.

DeBenedictis, E. P., Badaroglu, M., Chen, A., Conte, T. M. & Gargini, P. (2017),' Sustaining Moore's law with 3D chips'. *Computer*, 50(8), pp. 69-73.

Dekkers, T., van der Woude, M. & Koulish, R. (2019),'Objectivity and accountability in migration control using risk assessment tools'. *European Journal of Criminology*, 16(2), pp.237-254.

Diamandis, P.H. & Kotler, S. (2012), *Abundance: The future is better than you think*. New York: Free Press.

Dormehl, L. (2015), *The Formula: How Algorithms Solve All Our Problems and Create More*. New York: Penguin.

Dorrestijn, S. (2012),'Theories and figures of technical mediation', in Gunn, W. & Donovan, J. (eds.), *Design and anthropology* (pp.219-230), London: Routledge.

Dorrestijn, S. & Verbeek, P. P. (2013),'Technology, wellbeing, and freedom: The legacy of utopian design'. *International Journal of Design*, 7(3), pp.45-56.

Dotterway, K. A. (1992). *Systematic analysis of complex dynamic systems: the case of the USS Vincennes*. Naval Postgraduate school. Online via https://apps.dtic.mil/sti/pdfs/ADA260260.pdf (Accessed January 2024).

Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. (2006),'New public management is dead—long live digital-era governance'. *Journal of public administration research and theory*, 16(3), pp. 467-494.

Dutta, S. & Lavin, B. (2022), *The Network Readiness Index 2022*. Washington: Portulans Institute.

Economist (2017), *Obituary: Stanislav Petrov was declared to have died on September 18th*. Online via https://www.economist.com/obituary/2017/09/30/obituary-stanislav-petrov-was-declared-to-have-died-on-september-18th (Accessed January 2024).

Edelman Trust Baramoter (2022). Online via https://www.edelman.com/sites/g/files/aatuss191/files/2022-01/2022%20Edelman%20Trust%20Barometer%20FINAL_Jan25.pdf (Accessed January 2024).

Eidinow, E. (2007). *Oracles, curses, and risk among the ancient Greeks*. New York: Oxford University Press.

Eijkman, Q. (2014),'Digital Security Governance and Risk Anticipation: What About the Role of Security Officials in Privacy Protection'. *International Political Sociology*, 8(1), pp. 116-118.

Ellul, J. (1962), 'The technological order'. *Technology and culture*, 3(4), pp.394-421.

Ellul, J. (1964), *The Technological Society*. Translated by Wilkinson, J., New York: Vintage Books.

Ellul, J. (1990), *The Technological Bluff*. Translated by Bromiley, G., Grand Rapids: Eerdmans.

Elms, A. C. & Milgram, S. (1966),'Personality characteristics associated with obedience and defiance toward authoritative command'. *Journal of Experimental Research in Personality*, 1(4), pp.282-289.

Epley, N., Waytz, A. & Cacioppo, J. T. (2007),'On seeing human: a three-factor theory of anthropomorphism'. *Psychological review*, 114(4), pp.864-886.

Ericson, R. V. & Haggerty, K. D. (1997). *Policing the risk society*. Toronto: University of Toronto Press.

Eubanks, V. (2018), *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

European Parliament (2023), *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*. Online via https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai (Accessed January 2024).

Experts Exchange (2016), 'Infographic : Processing Power Compared. Online via https://www.datasciencecentral.com/processing-power-compared-infographics/ (Accessed January 2024).

Februari, M. (2023), *Doe zelf normaal: Menselijk recht in tijden van datasturing en natuurgeweld*. Amsterdam: Prometheus.

Feenberg, A. (1991). *Critical theory of technology*. New York: Oxford University Press.

FHWA (2024), *State motor vehicle registrations, by years, 1900-1995*. Online via https://www.fhwa.dot.gov/ohim/summary95/mv200.pdf (Accessed January 2024).

Fiske, S.T. & Taylor, S.E. (1994). *Social Cognition*. New York: McGraw-Hill.

Flathman, R. E. (1980). *The practice of political authority*. Chicago: University of Chicago Press.

Floridi, L. (2013), *The Ethics of Information*. Oxford: Oxford University Press.

Florence, J. (2006),'Making the no fly list fly: A due process model for terrorist Watchlists'. *The Yale Law Journal*, 115(1), pp. 2148-2181.

Fogg, B. J. & Nass, C. (1997),'Silicon sycophants: the effects of computers that flatter'. *International journal of human-computer studies*, 46(5), pp. 551-561.

Fogg, B. J. & Nass, C. (1997),'How users reciprocate to computers: an experiment that demonstrates behavior change'. *CHI'97 extended abstracts on Human factors in Computing Systems*, pp. 331-332.

Fogg, B. J. (2003), *Persuasive technology: using computers to change what we think and do*. San Francisco: Morgan Kaufmann Publishers.

Follow the Money (2023), *Dubieus algoritme van de politie 'voorspelt' wie in de toekomst geweld zal plegen*. Online via https://www.ftm.nl/artikelen/nederlandse-politie-gebruikt-minority-report-algoritme?share=nmsyzH6Tz/581UYGPAGzal4n79JX3JH4u81m9ze7F3iF5Fid2WLlMrxOUiXgUxw=&utm_medium=social&utm_campaign=sharebuttonleden&utm_source=whatsapp (Accessed January 2024).

Foucault, M. (1977), *Discipline and Punish*, London: Penguin.

French, S.E. & Lindsay, L.N. (2022),'Artificial Intelligence in Military Decision-Making: Avoiding Ethical and Strategic Perils with an Option-Generator Model', in Koch, B. & Schoonhoven, R. (eds.), *Emerging Military Technologies: Ethical and Legal Perspectives* (pp.53-74). Leiden: Brill Nijhoff.

Friedrich, C. J. (1972). *Tradition and authority*. Basingstoke: Macmillan.

Friedman, R. B. (1990),'On the concept of authority in political philosophy'. In Raz. J. (ed.), *Authority* (pp. 56-91). New York: New York University Press.

Friedman, B. (1995),'"It's the computer's fault" reasoning about computers as moral agents'. *Proceedings of the conference on Human factors in Computing Systems*, pp. 226-227.

Frissen, V. & van Lieshout, M. (2006),'ICT in everyday life: the role of the user'. In Verbeek, P.P. & Slob, A. (eds.), *User Behavior and Technology Development: Shaping Sustainable Relations Between Consumers and Technologies* (pp. 253-262). Dordrecht: Springer.

Fromm, E. (1936). Studies on Authority and the Family. Sociopsychological Dimensions. In Kassouf, S. (2020) (ed.) *Erich Fromm Early Writings* (pp.8-58). Online via https://www.fromm-gesellschaft.eu/index.php/en/publications/fromm-forum-english/671-fromm-forum-24-2020-en *(*Accessed January 2024).

Fry, H. (2018). *Hello World: Being Human in the Age of Algorithms*. London: Penguin.

Furedi, F. (2013). *Authority: a sociological history*. New York: Cambridge University Press.

Furedi, F. (2015),' Bringing Historical Dimensions Into the Study of Social Problems: The Social Construction of Authority'. *Qualitative Sociology Review* 11(2), pp.94-108.

Galinsky, K. (1996), *Augustan Culture: An Interpretive Introduction*. Princeton: Princeton University Press.

Garland, D. (2003),'The rise of risk'. In Ericson, R. (ed.) *Risk and morality* (pp. 48-86). Toronto: Toronto University Press.

Garthoff, R. L. (1994). *The great transition*. Washington: Brookings Institution Press.

Gaver, W.W. (1991),'Technology affordances'. *Proceedings of the conference on Human Factors in Computing Systems*, pp. 79–84.

Geiger, G. (2021),'How a discriminatory algorithm wrongly accused thousands of families of fraud'. *Vice*. Online via https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud (Accessed January 2024).

Geiskkovitch, D. Y., Cormier, D., Seo, S. H. & Young, J. E. (2016),'Please continue, we need more data: an exploration of obedience to robots. *Journal of Human-Robot Interaction*, *5*(1), pp.82-99.

General Data Protection Regulation (2016), Official Journal of the European Union. Online via https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 (Accessed January 2024).

Giddens, A. (1990), *The Consequences of Modernity*. Cambridge: Polity Press.

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Cambridge: Polity Press.

Giddens, A. (1994),'Living in a Post-Traditional Society', in Beck, U., Giddens, A. & Lash, S. (eds.), *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order* (pp. 56-109). Cambridge: Polity Press.

Giddens, A. (1998),'Risk Society: the Context of British Politics', in Franklin, J. (ed.), *The Politics of Risk Society* (pp. 23-35). Cambridge: Polity Press.

Giddens, A. (1999). Risk and responsibility. *The Modern Law Review*, 62(1), pp.1-10.

Giddens, A. (2002). *Runaway world: How globalization is reshaping our lives*. London: Profile Books Ltd.

Gillespie, N., Lockey, S., Curtis, C., Pool. J. & Akbari, A. (2023), *Trust in Artificial Intelligence: A Global Study*. The University of Queensland and KPMG Australia. Online via https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf (Accessed January 2024).

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011), *Heuristics: The foundations of adaptive behavior*. New York: Oxford University Press.

Gladden, M. (2014),'The Social Robot as 'Charismatic Leader': A Phenomenology of Human Submission to Nonhuman Power', *Sociable Robots and the future of Social Relations: proceedings of Robo-philosophy*, 273(1), pp.329-339.

Gladwell, M. (2008), Outliers: The story of success. London: Penguin.

Goffman, E. (1959), *The Presentation of Self in Everyday Life*. London: Penguin.

Goffman, E. (1967). Interaction ritual: essays on face-to-face interaction. New York: Pantheon Books.

Goffman, E. (1983),'The Interaction Order', *American Sociological Review*, 48(1), 1-17.

Goffman, E. (1986), *Frame analysis: An essay on the organization of experience*. Boston: Northeastern University Press.

Graesser, A. & Whitten, S. (2001),'Scripts of the Mind and Educational Reform', *Contemporary Psychology*, 46(3) pp. 261-262.

Graham, S. & Wood D. (2003),'Digitizing surveillance: Categorization, space, inequality'. *Critical Social Policy*, 23 (2), pp. 227-248.

Grant, M. (1969) *From Imperium to Auctoritas: A Historical Study of Aes coinage in the Roman Empire, 49 B.C. – A.D. 14*. Cambridge: Cambridge University Press.

Gray, M. (2003),'Urban surveillance and panopticism: will we recognize the facial recognition society?', *Surveillance & Society*, 1(3), pp. 314-330.

Greengard, S. (2015). *The Internet of Things*. Cambridge: MIT press.

Guzek, D. (2015),'Discovering the digital authority: Twitter as reporting tool for papal activities'. *Online: Heidelberg Journal of Religions on the Internet*, 9(1), pp.63-81.

Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M. (2013),'Internet of Things (IoT): A vision, architectural elements, and future directions'. *Future generation computer systems*, 29(7), pp.1645-1660.

Guthrie, S. E. (1993). *Faces in the clouds: A new theory of religion*. New York: Oxford University Press.

Hamilton, M. (2015),'Adventures in risk: predicting violent and sexual recidivism in sentencing law'. Arizona State Law Journal, 47(1), pp. 1-62.

Hardwig, J. (1986),'Epistemic Dependence'. *The Journal of Philosophy*, 82(7), pp.335-349.

Haring, K., Mosley, A., Pruznick, S., Fleming, J., Satterfield, K., de Visser, E., Tossel, C. & Funke, G. (2019),'Robot authority in human-machine teams: effects of human-like appearance on compliance', *International Conference on Human-Computer Interaction*. Orlando: Springer, pp.63-78.

Hayes, P., van de Poel, I. & Steen, M. (2020),'Algorithms and values in justice and Security'. *AI & Society*, 35(1), pp. 533-555.

Heidegger, M. (1927). *Being and time: A translation of Sein und Zeit*. Translated by Stambaugh, J. (2010), New York: SUNY press.

Heylighen, F. (2002). *Complexity and Information Overload in Society: why increasing efficiency leads to decreasing control*. Online via http://pespmc1.vub.ac.be/Papers/Info-overload.pdf (Accessed January 2024).

Hilbert, M. (2012),'How much information is there in the "information society"?', *Significance*, *9*(4), pp.8-12.

Hilbert, M. & López, P. (2011),'The world's technological capacity to store, communicate, and compute information'. S*cience*, *332*(6025), pp.60-65.

Hilbert, M. & López, P. (2012a),'Info Capacity| How to Measure the World's Technological Capacity to Communicate, Store and Compute Information? Part I: Results and Scope'. *International Journal of Communication*, 6(1), pp.956-979.

Hilbert, M. & López, P. (2012b),'Info Capacity| How to Measure the World's Technological Capacity to Communicate, Store and Compute Information? Part II: Measurement Unit and Conclusions'. *International Journal of Communication*, 6(1), pp.936-955.

Hildebrandt, M. (2011),'Legal Protection by Design: Objections and Refutations'. *Legisprudence*, 5(2), pp. 223–248.

Hildebrandt, M. & Gutwirth, S. (2008), *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Heidelberg: Springer

Hill, K. (2020), *Wrongfully accused by an algorithm*, The New York Times, online via https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (Accessed January 2024).

Hoff, K. A. & Bashir, M. (2015),'Trust in automation: Integrating empirical evidence on factors that influence trust'. *Human factors*, 57(3), pp.407-434.

Hoffman, D. (2009). *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy*. New York: Doubleday.

Hofstede, G. (1980), Culture's consequences: International differences in work-related values. Newbury Park: Sage.

Hofstede, G., Hofstede, G.J. & Minkov, M. (2010), *Cultures and Organizations: Software of the mind. Intercultural cooperation and its importance for survival*. New York: McGraw Hill.

Hofstede, G. (2024), *Country Comparison Tool France*. Online via https://www.hofstede-insights.com/country-comparison-tool?countries=france (Accessed January 2024).

Hume, D. (1757), *The Natural History of Religion.* Online via https://oll.libertyfund.org/titles/robertson-the-natural-history-of-religion (Accessed January 2024).

Hurley, D. (2018),'Can an Algorithm Tell me When Kids are in Danger?', *The New York Times Magazine*, Online via https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html?smid=pl-share (Accessed January 2024).

IBM (2017), *IBM has proudly announced it managed to fit 30 million transistors into a space of a fingernail*. Online via https://www.sciencealert.com/new-computer-chips-can-fit-30-million-transistors-on-your-fingertip (Accessed January 2024).

Ihde, D. (1990), *Technology and the lifeworld: From garden to earth*. Bloomington: Indiana University Press.

International Telecommunications Union (2018), *Measuring the information Society Report*. Geneva: Switzerland. Online via https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2018/MISR-2018-Vol-1-E.pdf (Accessed January 2024).

Introna, L. & Nissenbaum, H. (2010), *Facial recognition technology a survey of policy and implementation issues*. Lancaster University Management School, Working Paper online via https://eprints.lancs.ac.uk/id/eprint/49012/1/Document.pdf (Accessed January 2024).

Ison, S. & Budd, L. (2016),'Parking', in Bliemer, M., Mulley, C. & Moutou, C. (eds.), *Handbook on Transport and Urban Planning in the Developed World* (pp.145-162). Cheltenham: Edward Elgar Publishing.

Jacobsen, G. (2011),'Class action filed over glitch jailing young people'. *The Sydney Morning Herald*, online via https://www.smh.com.au/technology/class-action-filed-over-glitch-wrongly-jailing-young-people-20110608-1fs1h.html (Accessed January 2024).

Jarvis, D. (2007),'Risk, globalisation and the state: A critical appraisal of Ulrich Beck and the world risk society thesis'. *Global society*, 21(1), pp.23-46.

Jarvis, J. (2009), *What would Google do?.* New York: HarperCollins.

Jensen, K. (1997),'Studying law - A risky business'. *Young*, 5(1), pp.21-38.

Joh, E. (2016), *Wrongful Arrest by Software*, Slate, online via https://slate.com/technology/2016/12/software-problems-are-leading-to-wrongful-arrests.html (Accessed January 2024).

Johnson, D. & Gardner, J. (2009),'Exploring mindlessness as an explanation for the media equation: a study of stereotyping in computer tutorials'. *Personal and Ubiquitous Computing*, 13(1), pp.151-163.

Johnson, D. & Wetmore, J. (2021), *Technology and society: Building our sociotechnical future*. Cambridge: MIT Press.

Jongepier, F. (2020),'Algoritmische autoriteit en epistemische onrechtvaardigheid', in De Ridder, J., Vliegenthart, R. & Zuure, J. (eds.), *Doen, durven of de waarheid?: Democratie in digitale tijden*. Amsterdam: Amsterdam University Press.

Jopson, D. (2010), We'll sort out court computer debacle, Coalition pledges. The Sydney Morning Herald. Online via https://www.smh.com.au/technology/well-sort-out-court-computer-debacle-coalition-pledges-20101228-199bx.html (Accessed January 2024).

JMIR (2024), *JMIR-e-collection "Chatbots and Conversational Agents*. Online via https://www.jmir.org/themes/763-chatbots-and-conversational-agents (Accessed January 2024).

Juma, C. (2016). *Innovation and its enemies: Why people resist new technologies*. New York: Oxford University Press.

Kahneman, D. & Tversky, A. (1973),'On the psychology of prediction'. *Psychological review*, 80(4), pp.237-251.

Kahneman, D. & Tversky, A. (1970),'Prospect Theory. An Analysis of Decision Making Under Risk'. *Econometrica*, 47(2), pp.263-292.

Kalman, I. (2015),*"Don't blame me, it's just the computer telling me to do this": Computer attribution and the discretionary authority of Canada Border Services Agency officers*. Halle/Salle: Max Planck Institute for Social Anthropology.

Kekes, J. (2003),'Pluralism and Moral Authority', in Chong, K. and Tan, S. and Ten, C. (eds.), *The Moral Circle and the Self: Chinese and Western Approaches* (pp.75-93). Chicago: Open Court.

Kelman, H. C. & Hamilton, V.L. (1989), *Crimes of Obedience*. New Haven: Yale Univeristy Press.

Keymolen, E. & Broeders, D. (2013),'Innocence lost: Care and control in Digital Youth Care'. *The British Journal of Social Work*, 43(1), pp.41-63.

Keymolen, E. (2016), Trust on the Line: a philosophical exploration of trust in the network era. PhD dissertation, Erasmus University Rotterdam. Online via https://repub.eur.nl/pub/93210 (Accessed January 2024).

Kilby, J. (2001),'Turning Potential into Realities: The Invention of the Integrated Circuit'. *ChemPhysChem*, *2*(8-9), pp.482-489.

Kilham, W. & Mann, L. (1974),'Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm'. *Journal of Personality and Social Psychology*, 29(5), pp. 696-702.

Kitchin, R. (2014), *The data revolution: Big data, open data, data infrastructures and their consequences*. Thousand Oaks: Sage.

Kleinrock, L. (2004),'The Internet rules of engagement: then and now'. *Technology in Society*, 26(2-3), pp.193-207.

Kockelkoren, P. (2007), Mediated Vision. Rotterdam: Veenman Publishers.

Kohli, A., Markowitz, P. L. & Chavez, L. (2011),'Secure communities by the numbers: An analysis of demographics and due process'. *Warren Institute of Law and Policy, UC Berkeley.* Online viahttps://indypendent.org/wp-content/uploads/2012/06/Secure_Communities_by_the_Numbers.pdf (Accessed January 2024).

Koops, B. J. (2011),'The (in) flexibility of techno-regulation and the case of purpose-binding'. *Legisprudence*, 5(2), pp. 171-194.

Krieger, L. (1977),'The idea of authority in the West'. *The American Historical Review*, 82(2), pp. 249-270.

Kravits, D. (2014), *How Obama officials Cried 'terrorism' to cover up a paperwork error.* Wired. Online via https://www.wired.com/2014/02/no-fly-coverup/ (Accessed January 2024).

Kruglanski, A. W., Raviv, A., Bar-Tal, D., Raviv, A., Sharvit, K., Ellis, S. & Mannetti, L. (2005),'Says who?: Epistemic authority effects in social judgment'. *Advances in experimental social psychology*, 37(1), pp. 345-392.

Kumar, K. & Thakur, G. S. M. (2012),'Advanced applications of neural networks and artificial intelligence: A review'. *International journal of information technology and computer science*, 4(6), 57-68.

Kuflik, A. (1999),'Computers in control: Rational transfer of authority or irresponsible abdication of autonomy?'. *Ethics and Information Technology*, 1(3), pp.173-184.

Kurzweil, R. (2005), *The singularity is near.* New York: Viking Press.

Ladd, J. (1989),'Computers and Moral Responsibility', in Gould, C. (ed.), The Information Web: Ethical and Social Implications of Computer Networking (pp.207-228). Boulder: Westview Press.

Langer, E. J., Blank, A. & Chanowitz, B. (1978),'The mindlessness of ostensibly thoughtful action: The role of" placebic" information in interpersonal interaction'. *Journal of personality and social psychology*, 36(6), pp. 635-642.

Langer, E. J. (1992),'Matters of mind: Mindfulness/mindlessness in perspective'. *Consciousness and cognition*, 1(3), pp. 289-305.

Langer, E.J (2000),'Mindful Learning'. *Current Directions in Psychological Science*, 9(6), pp.220-223.

Langer, E. J., & Moldoveanu, M. (2000),'The Construct of Mindfulness'. *Journal of social issues*, 56(1), pp. 1-9.

Langer, E., Pirson, M., & Delizonna, L. (2010). The mindlessness of social comparisons. *Psychology of Aesthetics, Creativity, and the Arts*, 4(2), pp. 68-74.

Langer, E. J. (2014), *Mindfulness (25th Anniversary edition).* Boston: Da Capo Press.

Lankton, N. K., McKnight, D. H. & Tripp, J. (2015),'Technology, humanness, and trust: Rethinking trust in technology'. *Journal of the Association for Information Systems*, 16(10), pp.880-918.

Lareki, A., de Morentin, J. I. M., Altuna, J. & Amenabar, N. (2017),'Teenagers' perception of risk behaviors regarding digital technologies'. *Computers in Human Behavior*, 68(1), pp. 395-402

Latour, B. (1986),' The powers of association'. In Law, J. (ed.), *Power, Action and Belief. A New Sociology of Knowledge?* London: Routledge and Kegan Paul.

Latour, B. (1992),'Where are the missing masses? The sociology of a few mundane artifacts', in Bijker, W. and Law, J. (eds.), *Shaping Technology/ Building Society: Studies in Socio-technical change* (pp. 225-258). Cambridge: MIT Press.

Lee, E. J., Nass, C. & Brave, S. (2000),'Can computer-generated speech have gender?: an experimental test of gender stereotype'. *Proceedings of the conference on Human factors in Computing Systems*. pp. 289-290.

Lee, J. E. R. & Nass, C. (2010),'Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication', in Latusek, D. & Gerbasi, A. (eds.), *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives* (pp. 1-15). Hersey: IGI Global.

Lee, K. M. & Nass, C. (2003),'Designing social presence of social actors in human computer interaction'. *Proceedings of the conference on Human Factors in Computing Systems*. pp. 289-296.

Lessig, L. (2006), *Code 2.0.* New York: Basic Books.

Levi, B. G., von Hippel, F. N. & Daugherty, W. H. (1987),'Civilian Casualities form" Limited" Nuclear Attacks on the USSR'. *International Security*, 12(3), pp. 168-189.

Levy, D. (2007), *Love and Sex with Robots: The evolution of human-robot relationships*. New York: HarperCollins.

Leenes, R. (2011),'Framing techno-regulation: An exploration of state and non-state regulation by technology'. *Legisprudence*, 5(2), pp.143-169.

Levenson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. *Computer*, 26(7), pp. 18-41.

Lincoln, B. (1994). *Authority: Construction and corrosion*. Chicago: University of Chicago Press.

Livingstone, S. (2009), *Children and the Internet*. Cambridge: Polity Press.

Lupton, D. (2013), *Risk (2nd edition)*. New York: Routledge.

Lupton, D. (2016),'Digital risk society', in Burges, A. & Alemanno, A. & Zinn, J. (eds.), *Routledge Handbook of Risk studies* (pp. 301-309). New York: Routledge.

Lupton, D. & Jutel, A. (2015),'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps'. *Social Science & Medicine*, 133(1), pp. 128-135.

Lupton, D. & Tulloch, J. (1998),'The adolescent 'unfinished body', reflexivity and HIV/AIDS risk'. *Body & Society*, 4(2), pp.19-34.

Lustig, C. & Nardi, B. (2015),'Algorithmic authority: The case of Bitcoin'. *International Conference on System Sciences*. pp. 743-752.

Lustig, C., Pine, K., Nardi, B., Irani, L., Lee, M. K., Nafus, D., & Sandvig, C. (2016),' Algorithmic authority: the ethics, politics, and economics of algorithms that interpret, decide, and manage'. *Proceedings of the conference on Human Factors in Computing Systems*. pp. 1057-1062.

Machiavelli, N. (1513), *The Prince*. Translated by George Bull (2003), London: Penguin.

Madhavan, P., Wiegmann, D. A. & Lacson, F. C. (2006),'Automation failures on tasks easily performed by operators undermine trust in automated aids'. *Human factors*, 48(2), pp.241-256.

Mallard, G. (2016), *Bounded Rationality and Behavioural Economics*. New York: Routledge.

Mamalian, C.A. (2011),'State of the Science of Pretrial Risk Assessment'. Pretrial Justice Institute. Online via https://biblioteca.cejamericas.org/bitstream/handle/2015/5466/%5BPJI%5DStateoftheScienceofPretrialRiskAssessment%282011%29.pdf?sequence=1&isAllowed=y (Accessed January 2024).

Manzey, D., Reichenbach, J. & Onnasch, L. (2012),'Human performance consequences of automated decision aids: The impact of degree of automation and system experience'. *Journal of Cognitive Engineering and Decision Making*, 6(1), pp.57-87.

Marakas, G.M. (1998), *Decision Support Systems in the 21st century*. Upper Saddle River: Prentice Hall.

Marshall, W. E. (2014),'On-street parking', in Ison, S. & Mulley, C. (eds.), *Parking: Issues and Policies* (pp.361-381). Bingley: Emerald.

Marx, G. (1995),'The engineering of social control: The search for the silver bullet', in Hagan, J. & Peterson (eds.), *Crime and inequality* (pp.225-46). Stanford: Stanford University Press.

Maslow, A. H. (1943),'A theory of human motivation'. *Psychological review*, 50(4), pp.370-396.

McBride, M., Carter, L. & Ntuen, C. (2012),'The impact of personality on nurses' bias towards automated decision aid acceptance'. *International Journal of Information Systems and Change Management*, 6(2), pp.132-146.

McGuire, W. J. (1961),'Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments'. *The Journal of Abnormal and Social Psychology*, 63(2), 326-332.

McLaughlin, P. (2007). *Anarchism and authority: A philosophical introduction to classical anarchism*. New York: Routledge.

McQuillan, D. (2016),'Algorithmic paranoia and the convivial alternative'. *Big Data & Society*, 3(2), pp.1-12.

Meeus, W. H. & Raaijmakers, Q. A. (1995),' Obedience in modern society: The Utrecht studies'. *Journal of Social Issues*, 51(3), pp. 155-175.

Meijer, A., Lorenz, L. & Wessels, M. (2021),'Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems'. *Public Administration Review*, 81(5), pp. 837-846.

Mendes, J. R. (2021),'Does the sustainability of the anthropocene technosphere imply an existential risk for our species? Thinking with Peter Haff'. *Social Sciences*, 10(8), pp.314-328.

Merleau-Ponty, M. (1962), *Phenomenology of Perception*. New York: The Humanities Press.

Merritt, S. M. & Ilgen, D. R. (2008),'Not all trust is created equal: Dispositional and history-based trust in human-automation interactions'. *Human Factors*, 50(2), pp. 194-210.

Merton, R. K. (1968). *Social theory and social structure*. New York: Simon and Schuster.

Meyer, E. (2024), *The Culture Map: begrijp hoe mensen denken, leidinggeven en dingen bereiken binnen verschillende culturen* (*10th edition*). Amsterdam: Business Contact.

Milakovich, M. E. (2012). *Digital governance: New technologies for improving public service and participation*. New York: Routledge.

Milgram, S. (1963),'Behavioral study of obedience'. *The Journal of abnormal and social psychology*, *67*(4), pp. 371-378.

Milgram, S. (1973),'The perils of obedience'. *Harper's Magazine*, 247(1483), pp. 62-78.

Milgram, S. (1974), *Obedience to Authority: An Experimental View (Modern thought edition, 2009)*. New York: HarperCollins.

Mill, J.S. (1832), *The collected works of John Stuart Mill, volume XXIII – newspaper writings august 1831 – October 1834 part II*, (eds.) Robson, A. & Robson, P. Toronto: University of Toronto Press.

Miller, G. A. (1956),'The magical number seven, plus or minus two: Some limits on our capacity for processing information'. *Psychological review*, 63(2), pp. 81-97.

Miller, A. G. (1990),'A perspective on Kelman and Hamilton's Crimes of obedience'. *Political Psychology,* 11(1), pp. 189–201.

Möllering, G. (2006), Trust: Reason, Routine, Reflexivity. Oxford: Elsevier.

Moon, Y. (2000),'Intimate exchanges: Using computers to elicit self-disclosure from consumers'. *Journal of consumer research*, 26(4), pp. 323-339.

Moon, Y. & Nass, C. (1996),'How "real" are computer personalities? Psychological responses to personality types in human-computer interaction'. *Communication research*, 23(6), pp. 651-674.

Moon, Y. & Nass, C. (1998),'Are computers scapegoats? Attributions of responsibility in human–computer interaction'. *International Journal of Human-Computer Studies*, 49(1), pp. 79-94.

Moore, G. E. (1965),'Cramming more components onto integrated circuits'. *Electronics*, 38(8), pp.114-117.

Mori, M. (1970). The uncanny valley. *Energy*, 7(4), pp. 33-35.

Morozov, E. (2013*), To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist*. London: Allen Lane.

Mosier, K. L. & Skitka, L. J. (1999),'Automation use and automation bias'. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(3), pp. 344-348.

Mosier, K. L. & Skitka, L. J. (2009),'Human decision makers and automated decision aids: Made for each other?', in Parasuraman, R. & Mouloua, M. (eds.), *Automation and human performance: Theory and application* (pp. 201–220). Boca Raton: CRC Press.

Mussweiler, T., Rüter, K. & Epstude, K. (2006),'The why, who, and how of social comparison: A social-cognition perspective', in Guimond, S., *Social comparison and social psychology: Understanding cognition, intergroup relations, and culture* (pp.33-54). New York: Cambridge University Press.

Nass, C. (2004),'Etiquette equality: exhibitions and expectations of computer politeness'. *Communications of the ACM*, 47(4), pp. 35-37.

Nass, C. & Gong, L. (1999),'Maximized modality or constrained consistency?'. *International Conference on Auditory-Visual Speech Processing*. Online via https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cf2c6ff5489392672228ddfb700^{7445347}de63 (Accessed January 2024).

Nass, C., & Lee, K. M. (2001),'Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction'. *Journal of experimental psychology: Applied*, *7*(3), pp. 171-181.

Nass, C. & Moon, Y. (2000),'Machines and mindlessness: Social responses to computers'. *Journal of social issues*, *56*(1), pp.81-103.

Nass, C., Moon, Y. & Green, N. (1997),'Are machines gender neutral? Gender-stereotypic responses to computers with voices'. *Journal of applied social psychology*, 27(10), pp. 864-876.

Nass, C., Moon, Y., Morkes, J., Kim, E. and Fogg, B.J. (1997),'Computers are Social Actors: A Review of Current Research', in Friedman, B. (ed.*), Human Values and the Design of Computer Technology* (pp. 137-162). Cambridge: Cambridge University Press.

Nass, C., Reeves, B. & Leshner, G. (1996),'Technology and roles: A tale of two TVs'. *Journal of Communication*, 46(2), pp.121-128.

Nass, C. & Steuer, J. (1993),'Voices, boxes, and sources of messages'. *Human Communication Research*, 19(4), pp. 504-527.

Nass, C., Steuer, J., Henriksen, L. & Dryer, D. C. (1994),'Machines, social attributions, and ethopoeia: Performance assessments of computers subsequent to "self-" or "other-" evaluations'. *International Journal of Human-Computer Studies*, 40(3), pp. 543-559.

Nass, C., Steuer, J., Tauber, E. & Reeder, H. (1993),'Anthropomorphism, agency, and ethopoeia: computers as social actors. *Proceedings of the conference on Human factors in computing systems*. pp. 111-112.

Nielsen, J. (2023*), Nielsen's law of internet bandwidth*. Online via https://www.nngroup.com/articles/law-of-bandwidth/ (Accessed January 2024).

Nissenbaum, H. (1994),'Computing and accountability'. *Communications of the ACM*, 37(1), pp. 72-81.

Noble, S.U. (2018), Algorithms of Oppression: How search engines reinforce racism. New York: New York University Press.

Noflylistkids (2024), Online via https://noflylistkids.ca/en/home/ (Accessed January 2024).

Nordhaus, W. D. (2007),' Two centuries of productivity growth in computing'. *The Journal of Economic History*, 67(1), pp.128-159.

Noorman, M. (2023), Computing and Moral Responsiblity. Online via https://plato.stanford.edu/entries/computing-responsibility/ (Accessed January 2024).

Norman, D. (2013). *The design of everyday things (Revised and expanded edition)*. New York: Basic Books.

NOS (2017), Parkeren in de stad: pas op voor de scanauto. Online via https://nos.nl/artikel/2156450-parkeren-in-de-stad-pas-op-voor-de-scanauto (Accessed January 2024).

NOS (2018), Pathé voor 19 miljoen euro opgelicht door nepmails 'hoofdkantoor'. Online via https://nos.nl/artikel/2258662-pathe-voor-19-miljoen-euro-opgelicht-door-nepmails-hoofdkantoor (Accessed January 2024).

NOS (2019), *Schietverenigingen waarschuwen voor computertest bij afgifte wapenvergunning*. Online via https://nos.nl/artikel/2304601-schietverenigingen-waarschuwen-voor-computertest-bij-afgifte-wapenvergunning (Accessed January 2024).

NOS (2021), *Problemen bij computersysteem dat uitkeringen regelt:'Gaat al twee jaar fout'*. Online via https://nos.nl/artikel/2383250-problemen-bij-computersysteem-dat-uitkeringen-regelt-gaat-al-twee-jaar-fout (Accessed January 2024).

Noyes, J., Cook, M., & Masakowski, Y. (2016), *Decision making in complex environments*. Abingdon: Routledge.

NRC Handelsblad (1988), *Verkoper eist legitimatie van koning*. Online via https://www.nrc.nl/nieuws/1988/12/23/verkoper-eist-legitimatie-van-koning-kb_000030016-a3606435 (Accessed January 2024).

NRC Handelsblad (2018), *Computer zegt Nee. Hoe Saskia twintig jaar vastliep in het systeem*. Online via https://www.nrc.nl/nieuws/2018/06/01/die-auto-staat-wel-op-uw-naam-goeiedag-a1605121/ (Accessed January 2024).

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.

Olson, J. E. (2003). *Data Quality: The Accuracy Dimension*. San Francisco: Morgan Kaufmann Publishers.

Packer, D. J. (2008),'Identifying systematic disobedience in Milgram's obedience experiments: A meta-analytic review'. *Perspectives on Psychological Science*, 3(4), pp. 301-304.

Pacholke, D. (2016), Resignation letter Dan Pacholke to Governer Inslee. Online via https://www.documentcloud.org/documents/2705087-Pacholke-Inslee-Letter.html (Accessed January 2024).

Padden, M. (2016), Corrections, water for Spokane Valley emerge as major topics. Online via https://mikepadden.src.wastateleg.org/corrections-water-for-spokane-valley-emerge-as-major-topics-2/ (Accessed January 2024).

Pak, R., Fink, N., Price, M., Bass, B. & Sturre, L. (2012),'Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults'. *Ergonomics*, 55(9), pp. 1059-1072.

Parasuraman, R. & Manzey, D. H. (2010),'Complacency and bias in human use of automation: An attentional integration'. *Human factors*, 52(3), pp. 381-410.

Parasuraman, R. & Miller, C. A. (2004),'Trust and etiquette in high-criticality automated systems'. *Communications of the ACM,* 47(4), pp. 51-55.

Parasuraman, R. & Riley, V. (1997),'Humans and automation: Use, misuse, disuse, Abuse'. *Human factors*, 39(2), pp. 230-253.

Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000),'A model for types and levels of human interaction with automation'. *Transactions on systems, man, and cybernetics - part A: Systems and Humans*, 30(3), pp.286-297.

Parkes, A. (2012),'Persuasive decision support: Improving reliance on decision aids'. *Pacific Asia Journal of the Association for Information Systems*, 4(3), pp. 1-13.

Parlementaire Ondervragingscommissie Kinderopvangstoeslag (2020), *Ongekend onrecht*. Online via https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf (Accessed January 2024).

Peeters, R. & Schuilenburg, M. (2018),'Machine Justice: Governing security through the bureaucracy of algorithms'. Information Polity 23(1), pp.267-280.

Pérez-Perdomono, R. & Friedman, L. (2003),'Legal cultures in the age of globalization', in Friedman, L. & Pérez-Perdomono, R. (eds.), *Legal Culture in the age of globalization: Latin America and Latin Europe* (pp.1-20). Stanford: Stanford University Press.

Perinbanayagam, R. S. (1974),'The definition of the situation: An analysis of the ethnomethodological and dramaturgical view'. *Sociological Quarterly*, 15(4), pp. 521-54

Peters, R.S. (1958), 'Authority'. *Aristotelian Society, Supplementary Volumes*, 32(1), pp.207-224.

Pew Research Center (2019), *Internet/Broadband Fact Sheet*. Online via https://www.pewinternet.org/fact-sheet/internet-broadband/ (Accessed January 2024).

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT press.

Pichai, S. (2018), *Artificial intelligence could be our savior, according to the CEO of Google*. Online via https://www.weforum.org/agenda/2018/01/google-ceo-ai-will-be-bigger-than-electricity-or-fire/ (Accessed January 2024).

Pierson, R. (1994),'The Epistemic Authority of Expertise'. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1(1), pp. 398-405.

Pinizzotto, A. & Davis, E. (1992),'Cop Killers and Their Victims'. *FBI Law Enforcement Bulletin,* 61(12), 9-11.

Pinker, S. (2018), *Enlightenment Now: The case for reason, science, humanism and progress*. New York: Viking Press.

Postman, N. (1993), *Technopoly: The surrender of culture to technology*. New York: Vintage Books.

Prins, R. (2014), Safety First: How local processes of securitization have affected the position and role of Dutch mayors. PhD dissertation, Erasmus University Rotterdam. Online via https://repub.eur.nl/pub/76918/ (Accessed January 2024).

Pruitt, T. C. (2011), *Authority and the production of knowledge in archeology*. PhD dissertation, University of Cambridge. Online via https://www.repository.cam.ac.uk/bitstreams/631159ca-3bfd-47a0-ae73-e5bf39496d81/download (Accessed January 2024).

Rainie, L. & Wellman, B. (2012), *Networked: The new social operating system*. Cambridge: MIT Press.

Raz, J. (1990), *Authority*, New York: New York University Press.

Reason, J. (1990), *Human Error*. Cambridge: Cambridge University Press.

Reason, J. (1997). *Managing the risks of organizational accidents*. Abingdon: Routledge.

Redden, J., Denick, L. & Warne, H. (2020),'Datafied Child Welfare Services: unpacking politics, economics and power'. *Policy Studies*, 41(5), pp. 507-526.

Reeves, B., & Nass, C. I. (1996), *The media equation: How people treat computers, television, and new media like real people and places*. Stanford: CSLI Publications.

Reinsel, D. Gantz, J. & Rydning, J. (2017), Data Age 2025: The evolution of Data to

Life-Critical. IDC Whitepaper. Online via https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf (Accessed January 2024).

Reuters (2018), Wells Fargo faces tax credit probes, new problems with mortgage borrowers. Online via https://www.reuters.com/article/us-wells-fargo-housing/wells-fargo-faces-tax-credit-probes-new-problems-with-mortgage-borrowers-idUSKBN1KO2LE/ (Accessed January 2024).

Rice, S. & Keller, D. (2009),'Automation reliance under time pressure'. *Cognitive Technology*, 14(1), pp. 36-44.

Richmond, M. (2015),'The Etymology of Parking'. *Arnoldia*, 73(2), pp. 19-24.

Rienks, R. (2015), *Predictive Policing: Kansen voor een veiligere toekomst*. Politieacademie. Online via https://www.politieacademie.nl/kennisenonderzoek/kennis/mediatheek/PDF/90696.PDF (Accessed January 2024).

Rittel, H. & Webber, M. (1973),'Dilemmas in general theory of planning'. *Policy Sciences*, 4(1), pp.155-169.

Riva, G., Vatalaro, F., Davide, F. & Alcaniz, M. (2005), *Ambient intelligence: The Evolution of Technology, Communication and Cognition Towards the Future of Human-Computer Interaction*. Amsterdam: IOS press.

Robinette, P., Li, W., Allen, R., Howard, A. M. & Wagner, A. R. (2016),'Overtrust of robots in emergency evacuation scenarios'. *International Conference on Human-Robot Interaction*. pp. 101-108.

Rood, J. (2013), *Wat is er mis met gezag?*. Rotterdam: Lemniscaat.

Rooksby, E. (2009),'How to be a responsible slave: Managing the use of expert information Systems'. *Ethics and Information Technology*, 11(1), pp. 81-90.

Rosalia, C., Menges, R., Deckers, I. & Bartneck, C. (2005),'Cruelty towards Robots'. *Proceedings of the Robot Workshop - Designing Robot Applications for Everyday Use, Goteborg*. Online via https://www.researchgate.net/profile/Christoph-Bartneck/publication/200508195_Cruelty_towards_robots/links/5695^{74608}eab58^{9}50ac3/Cruelty-towards-robots.pdf (Accessed January 2024).

Rosling, H., Rosling, O. & Rosling- Rönnlund, A. (2019), *Factfulness: Ten reasons we're wrong about the world – and why things are better than you think*. London: Sceptre.

Ross J. M. (2008). Moderators of trust and reliance across multiple decision aids. PhD dissertation, University of Central Florida. Online via https://stars.library.ucf.edu/etd/3754/ (Accessed January 2024).

Roubroeks, M. A. (2014), *Understanding social responses to artificial agents: building blocks for persuasive technology*. PhD dissertation, Eindhoven University of Technology. Online via https://pure.tue.nl/ws/portalfiles/portal/3783290/774470.pdf (Accessed January 2024).

Russel, S. & Norvig, P. (2021), *Artificial Intelligence. A Modern Approach (4th edition)*. London: Pearson.

Sagarin, B. J., Cialdini, R. B., Rice, W. E. & Serna, S. B. (2002),'Dispelling the illusion of invulnerability: the motivations and mechanisms of resistance to persuasion'. *Journal of personality and social psychology*, 83(3), pp.526-541.

Sagarin, B. J., & Wood, S. E. (2007),'Resistance to influence', in Pratkanis, A. (ed.), *The science of social influence: Advances and future progress* (pp.321-340). New York: Psychology Press.

Salinger, M. (2010),'Behavioral economics, consumer protection, and antitrust'. *Competition Policy International*, 6(1), pp.65-86.

Salovey, P. & Mayer, J. D. (1990),'Emotional intelligence'. *Imagination, cognition and personality*, 9(3), pp.185-211.

Sanchez, J. Rogers, W. A., Fisk, A. D., & Rovira, E. (2014),'Understanding reliance on automation: effects of error type, error distribution, age and experience'. *Theoretical issues in ergonomics science*, 15(2), pp.134-160.

Sandberg, A. & Bostrom, N. (2006a),'Converging Cognitive Enhancements'. *Annals of the New York Academy of Sciences*, 1093(1), pp. 201-227.

Sandberg, A. & Bostrom, N. (2006b),'Cognitive enhancement: a review of technology'. Online via https://www.researchgate.net/profile/Anders-Sandberg-2/publication/240711096_Cognitive_Enhancement_A_Review_of_Technology/links/0c9605305e3cdec0f8000000/Cognitive-Enhancement-A-Review-of-Technology.pdf (Accessed January 2024).

Sauer, S., Walach, H., Schmidt, S., Hinterberger, T., Lynch, S., Büssing, A. & Kohls, N. (2013),'Assessment of mindfulness: Review on state of the art'. *Mindfulness*, 4(1), pp. 3-17.

Schakel, J. K., Rienks, R. & Ruissen, R. (2013),'Knowledge-Based Policing: Augmenting Reality with Respect for Privacy', in Custers, B., Schermer, B. & Zarsky, T. (eds.), *Discrimination and Privacy in the Information Society* (pp. 171-189). Heidelberg: Springer.

Schank, R. C. (1975),'Using knowledge to understand', in Schank, R.C. & Nash-Weber, B (eds.), *Theoretical issues in natural language processing* (pp. 117-121). Arlington: Tinlap Press.

Schank, R.C. (1995), *What we learn by doing*. Technical Report No. 60, Northwestern University, Institute for Learning Sciences. Online via https://web-archive.southampton.ac.uk/cogprints.org/637/1/LearnbyDoing_Schank.html?__hstc=245426011.6f51aad28a6d480b0896ec071bae-4c3d.1394496000020.1394496000021.1394496000022.1&__hssc=245426011.1.1394496000023&__hsfp=3077594379 (Accessed January 2024).

Schank, R.C. & Abelson, R.P. (1977), *Scripts, Plans, Goals and Understanding*. New Jersey: Lawrence Erlbaum Associates.

Scherer, M. U. (2016),'Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies'. *Harvard Journal of Law & Technology*, 29(2), pp. 353-400.

Schwarz, E. (2007),'Hybridity and humility: What of the human in posthuman security?', in Eroukmanhoff, C. & Harker, M. (eds.), *Reflections on the Posthuman in International Relations* (pp.29-39). Bristol: E-International Relations Publishing.

Scott, J. (2020), *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven: Yale University Press.

Searle, J. R. (1980),'The intentionality of intention and action'. *Cognitive science*, 4(1), pp.47-70.

Sennett, R. (1981), *Authority*. New York: Vintage Books.

Sharkey, J. (2008),'Not to Small to Appear on a Big No-Fly Watch List'. *The New York Times*. Online via https://www.nytimes.com/2008/09/30/business/30road.html (Accessed January 2024).

Shoup, D. (2011), *The high cost of free parking*. New York: Routledge.

Simon, H. (1972),'Theories of Bounded Rationality', in McGuire, C.B. and Radner, R. (eds.), *Decision and Organization* (pp.161-176). Elsevier, Amsterdam.

Simon, H. (1997), *Administrative Behavior: A Study of Decision-Making Processes In Administrative Organizations*. New York: The Free Press.

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. New York: Appleton-Century-Crofts.

Skitka, L. J., Mosier, K. L. & Burdick, M. (1999),'Does automation bias decision-making?. *International Journal of Human-Computer Studies*, 51(5), pp. 991-1006.

Skitka, L. J., Mosier, K. & Burdick, M. D. (2000),'Accountability and automation Bias'. *International Journal of Human-Computer Studies*, 52(4), pp.701-717.

Skitka, L. J., Mosier, K. L., Burdick, M. & Rosenblatt, B. (2000),'Automation bias and errors: are crews better than individuals?'. *The International journal of aviation psychology*, 10(1), pp.85-97.

Snapper, J. W. (1985). Responsibility for computer-based errors. *Metaphilosophy*, 16(4), pp. 289-295.

Snellen, I. (1998),'Street Level Bureaucracy in an Information Age', in Snellen, I. & Van de Donk, W. (eds.), *Public Administration in an Information Age. A Handbook* (pp.497-505). Amsterdam: IOS press.

Spain, R. D. & Madhavan, P. (2009),'The role of automation etiquette and pedigree in trust and dependence'. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53(4), pp. 339-343.

Spiegel (2010), *Der Mann, der den dritten Weltkrieg verhinderte*. Online via http://www.spiegel.de/einestages/vergessener-held-a-948852.html (Accessed January 2024).

Starr, S. B. (2015),'The risk assessment era: An overdue debate'. *Federal Sentencing Reporter*, 27(4), pp.205-206.

Stevenson, M. (2018),'Assessing risk assessment in action'. *Minnesota Law Review*, *103*, pp.303-384.

Suchman, L. (2007), Human-Machine Reconfigurations: Plans and Situated Actions. New York: Cambridge University Press.

Sundar, S. & Nass, C. (2000),'Source orientation in human-computer interaction: Programmer, networker, or independent social actor'. *Communication research*, 27(6), pp.683-703.

Supreme Court of Kentucky (2011), *Order Approving Judicial Guidelines for Pretrial Release and Monitored Conditional Release*. Online via https://kycourts.gov/courts/supreme/Rules_Procedures/201112.pdf (Accessed January 2023).

Susskind, J. (2018), *Future politics: Living together in a world transformed by tech*. Oxford: Oxford University Press.

Susskind, R. E., & Susskind, D. (2015), *The future of the professions: How technology will transform the work of human experts*. Oxford: Oxford University Press.

Swierstra, T. (1997),'From critique to responsibility: the ethical turn in the technology debate'. *Society for Philosophy and Technology*, 3(1), pp. 45-48.

Swierstra, T. (2013),'Nanotechnology and Technomoral Change'. *Ethica E Politica*, 15(1), pp.200-219.

Szalma, J. L. & Taylor, G. S. (2011),'Individual differences in response to automation: The five-factor model of personality'. *Journal of Experimental Psychology: Applied*, 17(2), pp.71-96.

Szaniawski, K. (1980),'Philosophy of decision making'. *Acta Psychologica*, 45(1), pp. 327-341.

Tang, Y.P., Shimizu, E., Dube, G.R., Rampon, C., Kerchner, G.A., Zhou, M., Liu, G. & Tsien, J.Z. (1999),' Genetic enhancement of learning and memory in mice'. *Nature*, 401(6748), pp.63-69.

Tenner, E. (1997). *Why things bite back: Technology and the Revenge of Unintended Consequences*. London: Fourth Estate.

Thaler, R. & Sunstein, R. (2008), *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.

The Daily Mail (2019), *British driver, 81, heading to see the Pope ends up a thousands miles away in Germany after typing 'Rom' not 'Rome' into his satnav*. Online via https://www.dailymail.co.uk/news/article-7075333/British-driver-81-heading-Rome-Pope-ends-GERMANY.html (Accessed January 2024).

The Daily Mail (2022), *Now that's eye-catching! Scientists develop a smartphone app that can detect signs of neurological conditions including Alzheimer's and ADHD based on selfie of your EYE*. Online via https://www.dailymail.co.uk/sciencetech/article-10777751/Smartphone-app-detect-signs-Alzheimers-ADHD-based-selfie-EYE.html (Accessed January 2024).

The Guardian (2023), *New artificial intelligence tool can accurately identify cancer*. Online via https://www.theguardian.com/society/2023/apr/30/artificial-intelligence-tool-identify-cancer-ai (Accessed January 2024).

The New York Times (1983), *Regan Denounces Ideology of Soviet as "Focus if Evil"*. Online via https://www.nytimes.com/1983/03/09/us/reagan-denounces-ideology-of-soviet-as-focus-of-evil.html (Accessed January 2024).

The New York Times (2006), *Jumping Through Hoops to Get Off The No-Fly List*, Online via https://www.nytimes.com/2006/02/14/business/jumping-through-hoops-to-get-off-the-nofly-list.html (Accessed January 2024).

The Telegraph (2007), *Hawk-eye system under the microscope*. Online via https://www.telegraph.co.uk/sport/tennis/wimbledon/2316731/Hawk-Eye-system-goes-under-the-microscope.html (Accessed January 2024).

The Wall Street Journal (2016), *Email Pioneer Ray Tomilson stayed Humble*. Online via https://www.wsj.com/articles/email-pioneer-ray-tomlinson-stayed-humble-1457605800 (Accessed January 2024).

The Washington Post (1999), *'I had a Funny Feeling in my Gut'*, Online via http://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter021099b.htm (Accessed January 2024).

The Washington Times (2008), *Air Marshals Grounded in List Mix-ups*. Online via https://www.washingtontimes.com/news/2008/apr/30/air-marshals-grounded-in-list-mix-ups/ (Accessed January 2024).

The Washington Post (2015), *How directions on the Waze app led to death in Brazil's favelas*. Online via https://www.washingtonpost.com/news/worldviews/wp/2015/10/05/how-directions-on-the-waze-app-led-to-death-in-brazils-favelas/ (Accessed January 2024).

The Washington Post (2022), *The Google Engineer who thinks the company's AI has come to life*. Online via https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/ (Accessed January 2024).

The Washington Post (2023), *AI Poses 'Risk of Extinction on par with Nukes', tech leaders warn*. Online via https://www.washingtonpost.com/business/2023/05/30/ai-poses-risk-extinction-industry-leaders-warn/ (Accessed January 2024).

Thomas, W. I. (1923). *The unadjusted girl*. Boston: Little Brown.

Thorndike, R.L. & S. Stein (1937),'An Evaluation of the Attempts to Measure Social Intelligence'. *Psychological Bulletin,* 34(5), pp. 275-285.

Tokmetzis, D. (2012), *De digitale schaduw: hoe het verlies van privacy en de opkomst van digitale profielen uw leven beïnvloeden*. Amsterdam: Spectrum.

Turkle, S. (2005), *The Second Self: Computers and the Human Spirit*. Cambridge: MIT Press.

Turing, A. (1948), *Intelligent Machinery*. Online via https://www.npl.co.uk/getattachment/about-us/History/Famous-faces/Alan-Turing/80916595-Intelligent-Machinery.pdf (Accessed January 2024).

Tversky, A. & Kahneman, D. (1973),'Availability: A heuristic for judging frequency and probability'. Cognitive Psychology, 5(2), pp.207–232.

United Nations (2019), *Brief by the United Nationals Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM c.s./De Staat der Nederlanden (Syri) before the District Court of The Hague (case number : C/09/550982/HA ZA 18/388).* Online via https://www.ohchr.org/sites/default/files/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf (Accessed January 2024).

United Nations (2023), *Technology and Innovation Report 2023. United Nations Conference on Trade and Development (UNCTAD)*. Online via https://unctad.org/tir2023 (Accessed January 2024).

Van Boeckel, J. (2015),'Never mind where, so long as it's fast: Jacques Ellul and La Technique'. *Technê*, 8(1), pp.214-229.

Van de Voort, M., Pieters, W. & Consoli, L. (2015),'Refining the ethics of computer-made decisions: a classification of moral mediation by ubiquitous machines'. *Ethics and Information Technology*, 17(1), pp. 41-56.

Van den Berg, B. (2009), *The situated self: Identity in a world of Ambient Intelligence*. PhD dissertation, Erasmus University Rotterdam. Online via https://repub.eur.nl/pub/15586 (Accessed January 2024).

Van den Berg, B. (2011),'Techno-elicitation: Regulating behaviour through the design of robots', in Van den Berg, B. & Klaming, L. (eds.), *Technologies on the stand: Legal and ethical questions in neuroscience and robotics* (pp.403-422). Nijmegen: Wolf Legal Publishers.

Van den Berg, B. & Leenes, R.E. (2013),'Abort, retry, fail: Scoping techno-regulation and other techno-effects', in Hildebrandt M, & Gakeer J. (eds.), *Human Law and Computer Law: Comparative Perspectives* (pp. 67-89). Heidelberg: Springer.

Van den Berg, B., & Keymolen, E. (2017),'Regulating security on the Internet: control versus trust'. *International Review of Law, Computers & Technology*, 31(2), pp.188-205.

Van den Hoven, M. J. (1998),'Moral responsibility, Public Office and Information Technology', in Snellen, I. & Van de Donk, W. (eds.), *Public Administration in an Information Age. A Handbook* (pp. 97-112). Amsterdam: IOS press.

Van der Linden, S., Leiserowitz, A., Rosenthal, S. & Maibach, E. (2017),'Inoculating the public against misinformation about climate change'. *Global Challenges*, 1(2), pp.1-7.

Van Eck, M. (2018), *Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming.* PhD dissertation, Tilburg University. Online via https://pure.uvt.nl/ws/portalfiles/portal/20399771/Van_Eck_Geautomatiseerde_ketenbesluiten.pdf (Accessed January 2024).

Vedder, A. & Klaming, L. (2010),'Human enhancement for the common good - Using neurotechnologies to improve eyewitness memory'. *AJOB Neuroscience*, 1(3), pp.22-33.

Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Pennsylvania: The Pennsylvania State University Press.

Verbeek, P. P. (2006a),'Materializing morality: Design ethics and technological mediation'. *Science, Technology, & Human Values*, 31(3), pp. 361-380.

Verbeek, P. P. (2006b). Persuasive Technology and Moral Responsibility Toward an ethical framework for persuasive technologies. *International conference on persuasive technology for human well-being.* pp. 1-15.

Verbeek, P.P. (2008),'Cyborg intentionality: Rethinking the phenomenology of human-technology relations'. *Phenomenology and the Cognitive Sciences,* 7(3), pp. 387–395.

Verbeek, P.P. (2011), *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: The University of Chicago Press.

Verbeek, P.P. (2015),'Beyond Interaction: a short introduction to mediation theory'. Interactions, 22(3), pp. 26-31.

Vitali-Rosati, M. (2018). *On editorialization: Structuring space and authority in the digital age*. Amsterdam: Institute of Network Cultures.

Volkova, M. & Pruel, N. (2019),'The leading role of expert communities in modern organizations'. *International Conference on Social, Economic, and Academic Leadership,* pp. 52-56.

Wachter, R. (2015). *The Digital Doctor: Hope, Hype and at the Dawn of Medicines Computer Age*. New York: Mcrgraw-Hill.

Wachter-Boettcher, S. (2017). *Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech*. New York: WW Norton & Company.

Waldrop, M. (2016),'The chips are down for Moore's law'. *Nature*, 530(7589), pp. 144-147.

Wang, L., Jamieson, G. A. & Hollands, J. G. (2009),'Trust and reliance on an automated combat identification system: The role of aid reliability and reliability disclosure'. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 51(3)*,* pp. 281–291.

Wasington Governor (2015), *Inslee announces immediate actions to fix 13-year-old sentencing errors discovered at state prisons*. Online via https://governor.wa.gov/news/2015/inslee-announces-immediate-actions-fix-13-year-old-sentencing-errors-discovered-state-prisons (Accessed January 2024).

Watt, E.D. (1982), *Authority*. London: Croom Helm.

Waytz, A., Cacioppo, J. & Epley, N. (2010),'Who sees human? The stability and importance of individual differences in anthropomorphism'. *Perspectives on Psychological Science*, 5(3), pp.219-232.

Waytz, A., Heafner, J. & Epley, N. (2014),'The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle'. *Journal of Experimental Social Psychology*, 52(1), pp. 113-117.

Weber, M. (1919),'Politics as Vocation', in Morgan, M. L. (ed.) (2011), *Classics of moral and political theory.* (pp. 1324-1360). Indianapolis: Hackett Publishing.

Weizenbaum, J. (1976), *Computer Power and Human Reason: From Judgement to Calculation*. New York: W.H. Freeman and Company.

Western, J, (2016),'What is *Global* in Security Studies?'. *Journal of global security studies*, 1(1), pp. 99-101.

Widlak, A. & Peters, R. (2018), *De digitale kooi: administratieve uitsluiting door informatiearchitectuur in de Basisregistratie Personen*. Den Haag: Boom Bestuurskunde.

Williamson, B. (2015),'Governing software: Networks, databases and algorithmic power in the digital governance of public education'. *Learning, Media and Technology*, 40(1), pp.83-105.

Wilson, P. (1983), *Second-hand knowledge: An inquiry into cognitive authority.* Westport: Greenwood Press.

Winner, L. (1978), Autonomous Technology: Technics-out-of-control as a Theme in Political Thought. Cambridge: MIT Press.

Winner, L. (1986), *The Whale and The Reactor: A Search for Limits in an Age of High Technology (2nd edition,* 2020). Chicago: The University of Chicago Press.

Winograd, T. & Flores, F. (1986), *Understanding Computers and Cognition: A new foundation for design*. New York: Ablex Publishing.

Wolff, R.P. (1990), 'The conflict between Authority and Autonomy', in Raz. J. (ed.), *Authority* (pp. 115-142). New York: New York University Press.

Wood, J. & Shearing, C. (2007), *Imaging Security*. Uffculme: Willan Publishing.

Woods, D. (2009),' Decomposing Automation: Apparent Simplicity, Real Complexity', in Parasuraman, R. & Mouloua, M. (eds.), *Automation and human performance: Theory and application* (pp. 3-19). Boca Raton: CRC Press

Workman, M. (2005),'Expert decision support system use, disuse, and misuse: a study using the theory of planned behavior'. *Computers in Human Behavior*, 21(2), pp. 211-231.

World Economic Forum (2018), *8 ways AI can help save the planet*. Online via https://www.weforum.org/agenda/2018/01/8-ways-ai-can-help-save-the-planet/ (Accessed January 2024).

Yudkowsky, E. (2023),'Pausing AI developments isn't enough. We need to shut it all down'. *Time Magazine*. Online via https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough (Accessed January 2024).

Zack, M. H. (2007),'The role of decision support systems in an indeterminate world'. *Decision Support Systems*, 43(4), pp. 1664-1674.

Završnik, A. (2021),'Algorithmic justice: Algorithms and big data in criminal justice settings'. *European Journal of criminology*, 18(5), pp. 623-642.

Zedner, L. (2003),'The concept of Security: an Agenda for Comparative Analysis'. *Legal Studies*, 23(1), pp. 153-175.

Zhang, C. & Lu, Y. (2021),'Study on artificial intelligence: The state of the art and future prospects'. *Journal of Industrial Information Integration*, 23(1), pp.1-9.

Zickuhr, K. (2013),'Who's not online and why'. *Pew Research Center*. Online via https://www.pewresearch.org/internet/2013/09/25/whos-not-online-and-why/ (Accessed January 2024).

Zuurmond, A. (1998),'From Bureaucracy to Infocracy; Are Democratic Institutions Lagging Behind?', in Snellen, I. & Van de Donk, W. (eds.), *Public Administration in an Information Age. A Handbook* (pp.259-271). Amsterdam: IOS press.

A ABOUT

C.V. DAAN WEGGEMANS

Education

Master Political Theory – Radboud University Nijmegen	2008 – 2010
Bachelor Political Science – Radboud University	2005 – 2008
Pre-Master Sociology – Radboud University	2006 – 2008
VWO – RSG Slingerbos | Levant - Harderwijk	1998 - 2005

Daan Weggemans (1986) studied Political Science at the Radboud University in Nijmegen (2005-2010).

From 2010-2016, he worked at Leiden University's Centre for Terrorism and Counterterrorism (CTC) and as a research fellow at the International Center for Counterterrorism (ICCT) in The Hague. During this period, he co-authored a book on the reintegration of extremists, worked on a databank on terrorist organisations with the National Coordinator for Counterterrorism (NCTV) and published several articles on contemporary security technologies and terrorism. He also served as an expert for the Dutch Court, the European Union, the Dutch Senate, and in various media outlets.

In 2016, he joined the Institute of Security and Global Affairs (ISGA) within the Faculty of Governance of Global and Global Affairs (FGGA) at Leiden University. Here he conducted his PhD research funded by the municipality of The Hague and was part of the Cyber Security Governance research group. At ISGA he also taught several interdisciplinary courses on contemporary security. During this time, he received research grants for projects on interdisciplinary education (NWO), families of foreign fighters (WODC) and intergenerational transmission of extremism (Ministry of Justice and Security). In 2018 he was appointed fellow at the National Support Center for Extremism (LSE).

From 2020-2024, he served as the director of the BA Security Studies and played a role in the design and development of the Bachelor Cybersecurity and Cybercrime. In 2023, he was appointed as a Leiden Teaching Fellow (LTA) and an expert member in commissions on knowledge security and monitoring of diversity and inclusion. He is also in the board of directors of the BA Cybersecurity and Cybercrime, and part of the editorial board for a book series on Digital Security.

KEY PUBLICATIONS

Broeders, D., Cristiano, F. & Weggemans, D. (2023),'Too close for comfort: cyber terrorism and information security across national policies and international diplomacy'. *Studies in Conflict & Terrorism*, 46(12), pp. 2426-2453.

Liem, M., & Weggemans, D. (2018),'Reintegration among high-profile ex-offenders'. *Journal of developmental and life-course criminology*, 4(4), pp. 473-490.

Weggemans, D.J. & De Graaf, B.A. (2016), *Reintegrating Jihadist Extremist Detainees: The challenge of helping former extremists back into society.* London: Routledge.

Weggemans, D. (2015),'Ik heb niets te verbergen'. VWO Staatsexamen Nederlands. *Centraal Instituut voor Toets Ontwikkeling* (CITO).

Weggemans, D.J., Bakker, E. & Grol, P. (2014),'Who are they and why do they go? The radicalization and preparatory processes of Dutch Jihadist foreign fighters'. *Perspectives on Terrorism*, 8(4), pp. 100-110.

Eijkman Q.A.M. & Weggemans D.J. (2013),'Open source intelligence and privacy dilemmas: Is it time to reassess state accountability?'. *Security and Human Rights,* 23(4), pp. 285-296.

Eijkman, Q.A.M. & Weggemans, D.J. (2011),'Visual Surveillance and the Prevention of Terrorism: What about the checks and balances?'. *International Review of Law, Computer & Technology*, 2(3), pp. 143-150.