



Universiteit
Leiden

The Netherlands

Computer says no: understanding digital authority

Weggemans, D.J.

Citation

Weggemans, D. J. (2025, February 13). *Computer says no: understanding digital authority*. Retrieved from <https://hdl.handle.net/1887/4180592>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4180592>

Note: To cite this publication please use the final published version (if applicable).

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 extent 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” (Mommson, 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 as 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 then 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 novel 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 & Koulisch, 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.

