Computer says no: understanding digital authority

Weggemans, D.J.

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.