Computer says no: understanding digital authority

Weggemans, D.J.

5 SITUATION

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

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

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

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

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

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

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

5.1 WE HAVE A SITUATION HERE!

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

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

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

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

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

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

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

5.2 DEFINITIONS, CUES AND SCRIPTS

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

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

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

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

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

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

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

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

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

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

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

5.3 AUTHORITY SITUATIONS AS ROLE CUES

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

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

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

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

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

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

5.4 ANTHROPOMORPHISM: SEEING HUMAN

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

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

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

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

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

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

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

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

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

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

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

5.5 THE MEDIA EQUATION: COMPUTERS ARE SOCIAL ACTORS

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

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

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

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

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

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

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

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

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

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

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

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

5.6 TECHNOLOGIES IN AUTHORITY SITUATIONS

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

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

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

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

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

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

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

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

5.7 A DIGITAL PROXY?

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

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

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

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

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

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

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

5