The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making

Rik Peeters

Centro de Investigación y Docencia Económicas (CIDE), División de Administración Pública, Carretera México-Toluca 3655, Colonia Lomas de Santa Fe, CP 01210, Alcaldía Cuajimalpa, Ciudad de México, México
E-mail: rik.peeters@cide.edu

Abstract. With the rise of computer algorithms in administrative decision-making, concerns are voiced about their lack of transparency and discretionary space for human decision-makers. However, calls to 'keep humans in the loop' may be moot points if we fail to understand how algorithms impact human decision-making and how algorithmic design impacts the practical possibilities for transparency and human discretion. Through a review of recent academic literature, three algorithmic design variables that determine the preconditions for human transparency and discretion and four main sources of variation in 'human-algorithm interaction' are identified. The article makes two contributions. First, the existing evidence is analysed and organized to demonstrate that, by working upon behavioural mechanisms of decision-making, the agency of algorithms extends beyond their computer code and can profoundly impact human behaviour and decision-making. Second, a research agenda for studying how computer algorithms affect administrative decision-making is proposed.

Keywords: Automated decision-making, algorithms, human-algorithm interaction, information architecture, administrative behaviour, transparency, discretion

Points for practitioners:

- The use of human oversight and override is crucial for fair and accountable algorithmic decision-making in public organizations.
- Designing transparency and discretion into algorithmic decision-making is not a sufficient guarantee that professionals in public organizations will also use the options for oversight and override in practice.
- The rise of machine learning and of information architectures complicate the actual possibilities of organizing algorithmic transparency and human discretion.
- It is key to take into account behavioural factors and organize the space for human decision-makers at all organizational levels (from data analysts to frontline workers) to make their own professional judgements rather than follow the default of algorithmic decisions.
- Professionals, especially those working in direct relation with citizens and clients, should have sufficient training to supervise the algorithms they are working with and to be able to use their discretionary space.

1. Introduction[1]

As the use of computer algorithms in public administration is expanding, concerns about their ability to produce fair and accountable decisions are voiced in both public and academic debates. Most of these

---

debates focus on two specific elements of algorithmic decision-making. First, algorithms operate as black boxes that impede human oversight to account for possible validity problems or bias in the input data or decision outcomes (Harcourt, 2007; Janssen & Van den Hoven, 2015; Kroll et al., 2016; Pasquale, 2015). Algorithms may lack transparency because they are protected by proprietary laws (Mittelstadt et al., 2016), because they analyse amounts of data that humans cannot process (Burrell, 2016) or because the constant algorithmic modification in machine learning impedes human oversight (Danaher, 2016; Binns, 2018). A second major concern is the reduction of the discretionary space for human decision-makers to override algorithmic decisions (Citron & Pasquale, 2014; Zouridis et al., 2020). 'Keeping humans in the loop' (Zarsky, 2011) is seen as crucial to assure individual administrative justice, to override potential errors and to adapt decisions to specific circumstances, as is also common practice in classic 'analogue' decision-making through regulated street-level discretion (Binns 2019; Van Eck, 2018).

Two assumptions lie behind these concerns of transparency and discretion. First, transparent algorithms will allow humans to verify the data and trace the steps that led to a specific decision and, thereby, to identify possible errors, bias or validity problems in statistical models. Second, human discretion will allow the override of algorithmic decisions that cause undesirable outcomes in specific cases. However, the argumentative leap that designing in oversight and override will automatically lead to their actual use is problematic. So far, the literature on algorithmic transparency, fairness and accountability in the context of public administration is limited in terms of empirical evidence. There are strong indications from other fields, however, that the agency of algorithms extends beyond their computer codes and immediate calculations and, instead, also affects behavioural mechanisms of human decision-making. For instance, studies on predictive policing and criminal justice – two policy areas where algorithmic applications have been analysed more extensively – show that correctional workers, judges and police officers may fail to scrutinize algorithms or use their own personal judgement (Hannah-Moffat, 2013; Monahan & Skeem, 2016).

These and other studies on 'human-algorithm interaction' (Van Eijk, forthcoming) highlight the importance of moving beyond sterile discussions of transparency and discretion and looking at the actual use of algorithmic applications in all their variety. 'Keeping humans in the loop' (Zarsky, 2011) may be a moot point if we fail to understand how algorithms impact human decision-making and how the organizational embeddedness of algorithmic applications may limit the practical possibilities for transparency and use of human discretion. The insight that technology is not a neutral or a-political instrument is well-established in science and technology studies (e.g. Winner, 1980; Latour, 1987). However, studies on algorithmic applications in public policy and administration have, so far, paid relatively little attention to the agency of algorithms. This article makes two contributions. First, the existing literature on the factors that influence algorithmic agency is reviewed and, second, a research agenda for studying how algorithms affect administrative decision-making is proposed.

In the following, first, the problems of algorithmic transparency and discretion are introduced, followed by a short methodological note. Next, a discussion of the design variables that determine the preconditions for human transparency and discretion is presented. This is followed by an analysis of the existing evidence on human-algorithm interaction in the public sector. The article ends with a summary of the main findings and the formulation of avenues for future research.

2. Transparency and discretion in algorithmic decision-making

2.1. Algorithms as a rationalizing force

Algorithmic decision-making refers here to any outcome generated by a sequence of digital rules and

criteria without human interference (Introna & Wood, 2004; Le Sueur, 2015). In principle, a computer algorithm is nothing more than a finite process consisting of a series of precisely defined instructions existing in a digital universe of standardized codes and classifications – like a digital cooking recipe of sorts. However, algorithms are far from trivial in their consequences. Beyond public sector research, the impact of algorithms has been extensively analysed and identified as a core part of contemporary business models (Zuboff, 2015) and as a driver for new forms of behavioural control (Couldry & Mejias, 2019). In public administration and public policy studies, a rapidly expanding body of literature has analysed, among other things, the application of algorithms in predictive policing (Bennett Moses & Chan, 2018; Smith & O'Malley, 2017), probation and sentencing decisions (Goel et al., 2016; Hamilton, 2015), the allocation of regulatory oversight resources (Yeung, 2018), risk assessment in child protection services (Gillingham, 2016) and individual administrative decisions (Van Eck, 2018). Even though much empirical work remains to be done, the literature suggests that the use of algorithmic decision-making is causing profound changes in the way governments deliver services, determine access to rights and benefits, allocate resources and enforce the law. In many ways, algorithms function as a "rationalizing force" (Pasquale, 2015: 15). More than just an instrument of government, the use of algorithms implies a different way of governing – of algorithmic governance or 'algocracy' (Aneesh, 2006; Danaher, 2016; Engin & Treleaven, 2019).

The use of algorithms promises more efficiency, less costs and higher quality in decision-making (Young et al., 2019: 306). Assessing whether algorithms actually deliver on these points goes beyond the scope of this article. Instead, the focus here is on the concerns voiced in the literature regarding the compatibility of algorithmic decision-making with principles of just and fair administration (Binns, 2019; Simmons, 2019). For example, studies in policing and criminal justice have argued that algorithmic decisions tend to reproduce bias towards already over-policed areas and target groups (Hannah-Moffat, 2016; Harcourt, 2007; Van Eijk, 2017). Elsewhere, studies have demonstrated that algorithmic decision-making can complicate an organization's ability to provide fair assessment of individual cases (Peeters & Widlak, 2018; Van Eck, 2018). Furthermore, accountability can be complicated because of the opacity of algorithmic decisions and the fact that algorithms do not leave a 'paper trail' like traditional 'analogue' decision-making procedures (Eubanks, 2018). In other words, the nature of computer algorithms has sparked two important concerns regarding the use of algorithmic decision-making in the public sector: lack of transparency and elimination of human discretion.

2.2. Algorithmic transparency

Computer algorithms can lack the transparency required for fairness in decision-making: if a decision-making procedure and the input data for that procedure are unknown, it is impossible to determine whether the actual outcome is fair and in accordance with legal requirements of due process (Pasquale, 2015; Ponce, 2005) or if an algorithmic risk assessment is based on a valid theoretical model and unbiased assumptions (Smith et al., 2017). Two underlying issues can be discerned. Algorithms may lack transparency in a very practical sense because they are protected by proprietary laws, which has been identified as a concern in the use of algorithms for sentencing and probation decisions (Pasquale, 2015). Algorithms may also lack transparency because they analyse amounts of data that humans cannot process (Kitchin, 2014) or because machine learning causes constant algorithmic modification that impedes human oversight (Binns, 2018; Eubanks, 2018). Algorithmic mechanics may be so complex – as in the case of data mining or machine learning – that they transcend human comprehension. Algorithms, then, are a rationalizing force that goes beyond human reason. In other words, transparency may be

compromised by either a lack of 'reviewability' of an algorithm (Danaher, 2016), or by epistemological limitations that inhibit the "reduction of algorithms to a human language explanation" (Zarsky, 2011: 293). A relevant issue here is whether transparency should concern the source code of an algorithm or should take the form of a right to a 'meaningful explanation' about the workings of an algorithm (Edwards & Veale, 2017).

2.3. Automated discretion

The reduction of human discretion in algorithmic decision-making may complicate administrative accountability. Regulated discretion in the application of universal laws in specific cases is considered a key mechanism in the Western legal tradition to ensure fair and proportional administrative decisions (Mashaw, 2007; Ostrom, 1996). Back in 2002, Bovens and Zouridis already demonstrated how automation replaces traditional bureaucracies with a system-level bureaucracy that generates mass amounts of decisions through automated procedures rather than human assessment – in the process, converting the traditional street-level bureaucrat (Lipsky, 1981) into a screen-level bureaucrat (cf. Landsbergen, 2004). When 'digital discretion' (Busch & Henriksen, 2018), 'artificial discretion' (Young et al., 2019) or 'automated discretion' (Zouridis et al., 2020) replaces human discretion, there is a risk that organizations hide their responsibility for individual decisions behind technological arguments, such as algorithmic complexity, system design flaws or lack of access to data (Fosch-Villaronga, 2019; Widlak & Peeters, 2018; Zalnieriute et al., 2019). Furthermore, a human eye is also deemed necessary to allow for tailor-made solutions or exceptions to prevent disproportional negative outcomes in individual cases (Peeters & Widlak, 2018).

3. Understanding algorithmic agency

3.1. The agency of algorithms

There have, so far, been relatively few empirical studies on the impact of automation on administrative decision-making in public administration (Bullock, 2019; Young et al., 2019: 303). There is, of course, a large body of work on e-government and how information and communication technology transforms service delivery and triggers organizational change (e.g. Bovens & Zouridis, 2002; Busch & Henriksen, 2018; Cordella & Tempini, 2015; Dunleavy et al., 2006; Fountain, 2004; La Porte et al., 2002; Margetts, 1999; Zuurmond, 1994). These studies encompass a large variety of ICT tools, including traditional "passive vehicles for the generation, transmission, and storage of digital data" (Young et al., 2019: 302). However, specifically regarding the impact of algorithms, concerns about transparency and discretion are often derived from the formal nature of algorithms and less from studies on their actual use in concrete organizational settings. Moreover, the (implicit) assumption that designing in the possibility for oversight and override into algorithms will lead to fair and accountable decision-making is still largely untested. There is evidence from related fields of research, however, that algorithms not only produce automated outcomes, but also impact human decision-making. For instance, studies on predictive policing and risk assessment tools in criminal justice have demonstrated that professionals may conflate correlation with causation (Hannah-Moffat, 2013) or risk with blame (Monahan & Skeem, 2016) and, thereby, misinterpret algorithms or attach too much weight to them.

This type of insights is well-established in science and technology studies and the sociology of technology. While it goes beyond the scope of this article to revise the literature in full, one relevant

issue from this field is to what extent and in what way technologies shape human behaviour and society. In 1980, Winner famously identified two ways in which artefacts can have politics – either through a deliberate technological design intended to change or enforce specific power relations or bias (such as defensive architecture) or through the power inherent to a technology from which functions and uses automatically follow (such as Postman's (1985) idea that visual imagery by default reduces everything to entertainment). The idea of technology as possessing agency is also central to actor-network theory, which holds that both human and non-human actors exist as part of broader networks in which they relate to each other (Latour, 1987). Rather than asking whether human design or technology itself produces power, agency is assigned to artefacts because humans relate to them in specific ways. Artefacts provide a 'script' for action which humans may follow up on (Akrich & Latour, 1992) – like a traffic light tells us to stop and a speed bump tells us to slow down.

In similar ways, it can be assumed that algorithms have agency. First, they might be designed with a specific bias or purpose in mind (such as profiling) or might determine human behaviour (such as excluding human interference). Second, algorithms may also impact human decision-making even if the possibility of oversight and override is designed into the algorithm. Algorithms provide 'actionable insights' (Ekbia et al., 2015). They are scripted and imply an action imperative. Matzner (2017) explicitly states that opening the black box of algorithms is not enough to gain understanding of their impact. Instead, the question should be: "What do algorithms do to subjects?" (ibid.: 28; cf. Schuilenburg & Peeters, forthcoming). Do humans follow or deviate from algorithmic outcomes? Under what circumstances? In other words, crucial in studying human-algorithm interaction is the age-old question 'how do people make decisions?' (Simon, 1947; cf. Bullock, 2019: 752).[2]

3.2. Methodological note

A literature review was carried out to get a better grasp of the existing empirical evidence on the agency of algorithms in public administration and public policy research. Relevant literature (academic articles and books) was selected in two steps (cf. Cooper, 2010; Tummers et al., 2015). First, a search was conducted in March 2020 for articles in the social science databases Web of Science and Scopus for mentions in title and keywords of 'automated decision-making', 'algorithms', 'automation', 'algorithmic transparency' and 'digital discretion'. The search was limited to public administration and public policy journals, both general and specialized journals on information technology, surveillance studies and frontline work – following the expectation that relevant evidence could be found there. Second, a forward and backward search (cf. Busch & Henriksen, 2018) was conducted to identify relevant publications outside the direct realm of public administration in the reference list of the selected articles and in publications that cite the selected articles.

Publications were included if they met the following criteria: 1) covering an algorithmic application in the public sector, 2) book publications, research reports or publications in double blind peer reviewed journals, 3) English or Dutch language publications and 4) published in the last ten years (with exceptions for several older key publications in the field). Following these criteria, a total of 63 publications was selected for analysis. Because research on e-government and algorithms is highly dispersed and specialized, the overview presented here does not claim to be exhaustive. Moreover, excluded are technical

---

[2]Outside the direct realm of public policy and public administration, the human side of decision-making with technology has been more extensively studied (e.g. Berendt & Preibusch, 2017; Zerilli et al., 2019). This article takes notice of the main insights from these studies, but focuses specifically on algorithmic applications in public administration.

discussions of algorithmic design, detailed legal discussions and algorithmic applications in the private sector. The selected publications were analysed by coding text segments that refer to the agency of algorithms beyond their computer codes through 1) their technological or organisational design or 2) the way human decision-makers use algorithmic applications. This follows well-established notions of technological agency in science and technology studies, as discussed above.

4. Algorithmic agency through design

Algorithmic applications may have agency designed into them (cf. Winner, 1980). Based on public administration literature, three key design variables are identified in the following: a) level of automation, b) type of algorithm (predictions or process automation) and c) organizational scope. Issues of human-algorithm interaction can play out in every algorithmic application, but these design variables influence the opportunities humans have to exercise oversight and override in the first place (cf. Bullock, 2019: 757). The following findings focus on the design characteristics of algorithms in organizational contexts in general and, thereby, exclude more specific objectives of algorithmic design such as 'nudging algorithms' (O'Keeffe et al., 2019; Ranchordás, 2019) or 'dark patterns' (Brignull & Rogers, 2003; Gray et al., 2018).

4.1. Human oversight and override as a continuum

The very nature of algorithms suggests the full exclusion of human agency. However, both in input and output some form of human decision-making is usually required. Above all, all algorithms are designed by humans. Even the most autonomous machine learning algorithms have an 'objective function' designed into them by a human being, which tells them what to search for. Often, the scope of data input is determined by humans as well. Human data- and performance-analysis is required to ensure the proper functioning of the algorithm. And in many cases, human follow-up decisions at operational organizational levels are needed to give algorithmic outcomes practical effects. In short, levels of automation vary, and, consequently, varying levels of transparency and discretion for human decision-makers at 'street-level' and 'screen-level' (Bovens & Zouridis, 2002) can be designed into algorithms.

For instance, in their analysis of military drones, Citron and Pasquale (2014) distinguish between 1) human-in-the-loop algorithms, 2) human-on-the-loop algorithms and 3) human-out-of-the-loop algorithms. Applied in the context of public administration, one can think of a tax office that uses algorithms to 1) merely select targets for auditing, 2) select targets and suggest enforcement interventions with human agents left to decide whether to follow these decisions or not, or 3) select targets and enforcement interventions autonomously without human intervention (Danaher, 2016). The 'absence' of human agency in algorithmic decision-making can, therefore, best be understood as a continuum (Young et al., 2019: 304, 306).

4.2. Decisions and predictions

Generally speaking, two types of algorithmic applications can be discerned in public organizations: decisions and predictions. First, algorithmic administrative decisions determine an individual citizen's status as eligible for rights (such as benefits, welfare state services) or obligations (such as taxation). These algorithms classify and categorize (Bowker & Star, 2000). The automation of primary processes is most common in routinized non-complex administrative tasks (Bovens & Zouridis, 2002), although it is argued that technological developments increasingly allow for automation in more complex and

non-routine frontline work as well, such as teaching, nursing and policing (Busch & Henriksen, 2018: 20). Generally, this type of automation leads to a transfer of discretion and transparency to the system-level were automation is designed and managed, thereby transforming classic street-level bureaucrats into screen-level bureaucrats (Bovens & Zouridis, 2002) or even eliminating them completely from standard procedures (Scholta et al., 2019).

Second, algorithmic prediction uses statistical analysis to profile individuals from a broader group based on specific characteristics or behavioural patterns, often with the purpose of determining a heightened risk (of, for instance, fraud or recidivism).[3] This form of algorithmic application is more common for non-routine tasks, such as informing decisions on allocation of police resources (Bennett Moses & Chan, 2018; Smith & O'Malley, 2017), risk assessment in child protection services (Gillingham, 2016) or sentencing and probation decisions (Goel et al., 2016; Hamilton, 2015), although it may also be applied for risk assessments on large data sets, such as the identification of possible tax fraud (Danaher, 2016). Predictive algorithms, by themselves, do not entail an individual administrative decision, but often serve to inform public officials (such as courts, tax offices, regulators) in charge of making those decisions (Houser & Sanders, 2017: 13). This leaves, at least formally, room for system- or street-level discretion. However, oversight is complicated given the large amounts of data that algorithms process and the complex pattern analyses they perform. Moreover, humans can predetermine the patterns that algorithms search for, but they can also allow algorithms to find patterns themselves (Kitchin, 2014; Zarsky, 2011: 291-292). Through machine learning, algorithms 'learn' new ways of classification and identification of patterns or anomalies (Aradau & Blanke, 2017; Binns, 2018), which indicates a shift in discretion from system designers to the algorithm itself (Hannah-Moffat, 2018).

4.3. The organizational scope of algorithms

Algorithms can be applied at various organizational levels. For instance, Young and others (2019: 305) distinguish between the individual, organizational and institutional level, referring to automation of mere individual officials' tasks, entire organizational processes or even multi-organizational policy formulation and goal setting. Elsewhere, Peeters and Widlak (2018) demonstrate how multiple government organizations can be held together by a single supra-organizational 'information architecture'. Whereas Bovens and Zouridis (2002) still defined their 'system-level bureaucracy' in terms of a single organization that 'owned' the entire production process, information architectures are characterized by a separation between organizations that gather, process and share data on the one hand and organizations that make the actual administrative decisions based on that data on the other hand. This both expands and fragments the system-level bureaucracy.

Within information architectures, a further distinction can be made between automated chain decisions and automated network decisions (Widlak et al., forthcoming). A 'chain' involves several hierarchically independent organizations that cooperate in a pre-defined sequential process towards a collective result. An example of this is data sharing between police and public prosecutor. Chain decisions are usually organized within a shared legal framework and with harmonized data definitions (Van Eck, 2018; Zouridis et al., 2020). In network decisions, however, there is no sequential process, no harmonization of definitions and no feedback mechanisms regarding data or administrative decisions. Albeit in differing ways, both

---

[3]There are also other forms of algorithmic prediction that can affect citizens indirectly, such as predictive modelling of flooding or traffic jams. Even though these models might lead to decisions that affect citizens further down the line, this type of algorithm is not further analyzed here since it does not directly imply a state-citizen interaction.

Table 1
Consequences of design variables for algorithmic transparency and discretion

	Transparency	Discretion
Level of automation		
Humans in the loop	Human oversight of target selection possible, transparency of intervention mechanism	Human override of target selection possible, human follow-up for intervention mechanism required
Humans on the loop	Human oversight of target selection and default intervention mechanism possible	Human override of target selection and default intervention mechanism possible
Humans out of the loop	Human oversight of target selection and default intervention mechanism possible, but unlikely	Human override of target selection and default intervention mechanism impossible
Type of algorithm		
Administrative decisions	Transparency high at system-level, low at screen-level (for routinized tasks) and varied at street-level (depending on level of automation)	Discretion high at system-level, low at screen-level (for routinized tasks) and varied at street-level (depending on level of automation)
Administrative prediction	Transparency low at street- and screen-level and varied at system-level (big data and machine learning algorithms cause interpretability issues)	Discretion generally high at street-level (non-routine tasks), low at screen-level and varied at system-level (e.g. discretionary shift to algorithm in machine learning)
Organizational scope		
Automated chain decisions	Data and decision transparency varied for both data owners and data users (depending on level of coordination and feedback mechanisms)	Varied for both data users and data owners (depending on level of automation)
Automated network decisions	Data transparency low for data users, varied for data owners (e.g. depending on machine learning); decision transparency low for data owners and varied for data users (depending on level of automation)	Varied for data owners (usually high in case of algorithmic prediction, low in case of data processing), usually low for data users

automated chain decisions and automated network decisions have a profound impact on the possibility of designing in algorithmic transparency and discretion. The separation of the ownership of data and the ownership of administrative decisions complicates algorithmic transparency for both data owners (regarding which administrative decisions are made with their data) and data users (regarding data origins and quality). Discretion is also affected since data users usually depend fully on the data provided by designated data owners (Peeters & Widlak, 2018).

In sum, the possibility for human decision-makers to have meaningful transparency and discretion varies according to the way algorithms are designed.

5. Algorithmic agency in human-algorithm interaction

5.1. Algorithms and human control problems

Existing evidence indicates that algorithms impact human decision-making even when the possibility of oversight and override is designed into the algorithm (cf. Latour, 1987). The findings of this section of the literature review apply, in principle, to administrative decision-making at all organizational levels. To better contextualize the limited amount of evidence on human-algorithm interactions in public administration and public policy, insights from human-machine systems are used to link classic notions of 'administrative behaviour' (Simon, 1947; cf. Bullock, 2019: 752) with the idea of 'control problems'

(Bainbridge, 1983; Zerilli et al., 2019). Control problems refer to "the tendency of the human within a human-machine control loop to become complacent, over-reliant or unduly diffident when faced with the outputs of a reliable autonomous system" (ibid.: 555).

5.2. Bounded rationality

Human agents may misinterpret algorithms and their outcomes, thereby expressing well-documented bounded rationality in their decision-making (Simon, 1947). For instance, Hannah-Moffat (2013: 278) observes that even well-trained practitioners often conflate correlation with causation in the interpretation of probability scores of high-risk offenders. Similarly, Werth (2017) demonstrates how parole officers in California tend to view a high-risk offender as an inherently dangerous subject. And finally, Monahan and Skeem (2016) show how judges often conflate risk and blame in their sentencing based on algorithmic risk scores. In other words, human decision-makers are susceptible to 'interpretative slippages' (Van Eijk, forthcoming). Here, algorithms have agency because humans face cognitive limitations in interpreting outcomes and processes which impede the use of oversight and override.

In terms of control problems (Zerilli et al., 2019), bounded rationality reflects the notion that humans have an epistemic disadvantage compared to machines ('capacity problem'). Machines can process more information at a higher speed than even the best trained humans are capable of: "There is [...] no way in which the human operator can check in real-time that the computer is following its rules correctly. One can therefore only expect the operator to monitor the computer's decision at some meta-level, to decide whether the computer's decisions are 'acceptable'" (Bainbridge, 1983: 776). This is, in the context of public administration, especially relevant for algorithmic predictions, where the problem is not only the speed or amount of data, but also the complex statistical patterns that computers establish. Moreover, expert knowledge of algorithms is more likely to be found at system-level (data analysts and system designers) than at street- or screen-level.

Furthermore, bounded rationality can also be linked to the 'attention control problem' (Zerilli et al., 2019). As machines become more reliable and encompassing, it is more difficult for humans to maintain adequate levels of attention and engagement needed for supervision. Automation is "most dangerous when it behaves in a consistent and reliable manner most of the time" (Banks et al., 2018: 283). In the words of Bainbridge: "[...] it is impossible for even a highly motivated human being to maintain effective visual attention towards a source of information in which very little happens, for more than half an hour. This means that it is humanly impossible to carry out the basic function of monitoring for unlikely abnormalities" (Bainbridge, 1983: 776). In accordance, especially in the case of routinized tasks, human decision-makers at screen-level are likely to face difficulties in identifying abnormalities that warrant the use of their discretion.[4]

5.3. Satisficing behaviour

Algorithms may be used as a convenient default for human decision-making, thereby reflecting well-known mechanisms of satisficing behaviour (Simon, 1947). Several policing and criminal justice studies indicate that human actors tend to not use the discretion they have and, instead, follow algorithmic

---

[4]Bounded rationality can also be linked to the 'currency problem'. The more algorithmic outcomes are used as the default decision, the more human skills deteriorate because of lack of use (Bainbridge, 1983: 775–776). This implies that without use, humans become less likely to use their discretion and, instead, more likely to follow algorithmic procedures and outcomes.

outcomes, or even adapt their behaviour to fit algorithmic outcomes. For instance, a report by the AI Now Institute states that high risk assessment scores "change the sentencing outcome and can remove probation from the menu of sentencing options the judge is willing to consider" (AI Now Institute, 2018: 13). A French parliamentary report on artificial intelligence draws a similar conclusion and states that "it is far easier for a judge to follow the recommendations of an algorithm [. . .] than to look at the details of the prisoner's record" and that it "is easier for a police officer to follow a patrol route dictated by an algorithm than to object to it" (Villani, 2018: 124). Otherwise, they would have to "defend their 'discretionary' decisions" (ibid.). In child abuse prevention, Eubanks (2018) describes the use of algorithms by caseworkers in a US county in their assessment whether to follow-up on calls in the local child welfare hotline. She finds that caseworkers are sometimes tempted to adjust their own professional risk assessment to the outcome produced by the algorithm.

In the context of automated network decisions, Peeters and Widlak (2018) find evidence for similar behavioural mechanisms in their case study of automated data sharing in the Dutch civil registry. First, data owners experience a bigger pressure to take on a role of enforcer because of the financial interests that data users have in keeping their client records up to date and free of fraud. Second, there are indications that technological efficiency or convenience may take precedence over legal guidelines of due process. Third, and finally, although data users are responsible for the administrative decisions they make regarding eligibility of citizens, they tend to rely completely on the data provided by the data owners. There is no incentive for a 'human eye' in their data control and screen-level decision-making. This is confirmed in another study on the way administrative errors spread across multiple organisations through automated data-sharing: data users take the data they receive at face value without verifying the data or looking at a citizen's factual situation (Widlak & Peeters, 2020).

These and other studies suggest that algorithms have agency because humans have an incentive to minimize their efforts of oversight and override. This is commonly seen as a form of strategic behaviour. It has no clear equivalent in the control problems identified by Bainbridge (1983), because these tend to focus on psychological mechanisms. However, it is clear that satisficing behaviour presents a control problem in itself since the result is the same: complacency and an underused potential of oversight and override.

5.4. Automation bias

Human agents may have a disproportional trust in the validity and rationality of algorithms, thereby producing an attitudinal bias in decision-making. There are indications that this cognitive mechanism is at play in the use of algorithms. In policing studies, this has been identified as a rationalization process and as a key characteristic of the 'techno-moral assemblage' (Werth, 2017: 822) of algorithmic decision-making. In social work research, there is evidence that caseworkers use less discretion when their algorithmic decision-making support tools are theory-based as compared to tools that require interpretation from the caseworkers themselves (Høybye-Mortensen, 2015). And a recent study on algorithm-assisted face recognition finds that it leads to a deterioration of human accuracy since human decision-makers are inclined to follow algorithmic outcomes as a default – thereby reducing the possibility of identifying false positives and possibly also of false negatives in face matching (Howard et al., 2020). In these and other examples, algorithms have agency because human operators see algorithms as rational, scientific and value-neutral (Silver, 2000).

This control problem has been identified as an attitudinal problem (Zerilli et al., 2019). As the quality of automation improves, humans become more complacent in their supervisory role. The operator "starts

Table 2
Consequences of human factors for algorithmic transparency and discretion

	Transparency	Discretion
Bounded rationality	Epistemic limitations to understanding algorithms at all levels, but especially likely among non-expert human agents at street- and screen-level (and even more so in the case of algorithmic prediction).	Misinterpretation of algorithmic outcomes at system-level or street-level; limited attention span to identify errors or validity problems at screen-level (for routinized tasks).
Satisficing behaviour	Human agents do not scrutinize algorithms on their validity because of organizational incentives of efficiency, accountability and (especially in the case of automated network decisions) task specialization.	Human agents use algorithmic outcomes as the default for decision-making because of organizational incentives of efficiency, accountability and (especially in the case of automated network decisions) task specialization.
Automation bias	Transparent algorithms are not scrutinized because of a belief in the scientific validity, neutrality and rationality of algorithmic procedures and outcomes.	Discretion is not used because of a belief in the scientific validity, neutrality and rationality of algorithmic procedures and outcomes.
Frontline work	Willingness to scrutinize algorithmic decisions in non-routine tasks because of professional standards or citizen orientation in frontline work.	Negotiated use of algorithms in non-routine tasks; trade-off with experiential, moral and affective forms of decision-making.

to assume that the system is infallible, and so will no longer actively monitor what is happening, meaning they have become complacent" (Pazouki et al., 2018: 299). Complacency may lead human operators to "trust the automated system so much that they ignore other sources of information, including their own senses" (ibid.). This is also linked to insights from cognitive dissonance theory (Festinger, 1957), prospect theory (Kahneman & Tversky, 1979) and the literature on technology acceptance (Davis, 1989). Studies have demonstrated that public officials are, relative to the general public, more likely to support information and communication technology and more confident in its ability to provide public services (Moon & Welch, 2005). In accordance, Moynihan and Lavertu (2012) found evidence for a technology preference among local election officials for e-voting applications based on a general faith in technology. Other have identified this bias in terms of cyberoptimism (Norris, 2001) or technological idolization (Goldfinch, 2007).

5.5. Counter-indications: Frontline work

In accordance with well-documented decision-making and coping mechanisms in frontline work (Lipsky, 1981; Maynard-Moody & Musheno, 2003; Tummers et al., 2015), there are also indications that human decision-makers at street-level resist the agency of algorithms. For instance, Binns (2019: 19) notes that human decision-makers exercise discretion in the use of algorithms according to their own convictions and commitments. Keddell (2019) finds that child protection professionals in New Zealand have reservations regarding working with predictive algorithms they are unable to explain to families, while at the same time they are being held accountable for decisions based on them. This can lead to what Elish (2019) calls 'moral crumple zones', where human actors bear the brunt of malfunctions in automated decision-making procedures they have little or no control over. Elsewhere, Hannah-Moffat and others (2009) observe that Canadian correctional workers deal with predictive algorithms on their own terms and sometimes overrule predictions and prefer to rely on their own expertise. Risk assessment in frontline work can, therefore, be best understood as a negotiated process in which different ways of assessing individuals – actuarial, experiential, moral and affective – are integrated (ibid.). This suggests that the outcomes of human-algorithm interactions highly depend on the specific organizational and professionals settings that allow human decision-makers to operate on behalf of citizens' interests rather than comply with algorithmic outcomes.

In sum, the factors identified above affect the way humans make use of oversight and override options built into algorithms.

6. Conclusion

Calls for algorithmic transparency (Pasquale, 2015) and 'keeping humans in the loop' (Zarsky, 2011) may be moot points if we fail to understand how algorithms impact human decision-making and how the design of algorithmic applications expands or limits the practical possibilities for transparency and use of human discretion. It is crucial to move beyond essentialist and functionalist approaches to algorithms and, instead, look at the actual use of algorithmic applications in all their variety and analyse the 'human-algorithm interaction' (Van Eijk, forthcoming). Through a review of the existing evidence on automated decision-making in public administration, two main sources of variation in the factual use of transparency and discretion have been identified:

1. Variety in algorithmic design: human agents are rarely fully 'out of the loop' and levels of oversight and override designed into algorithms should be understood as a continuum. However, transparency is likely higher in algorithmic administrative decisions than in algorithmic predictions; use of discretion at street-level is likely lower in algorithmic administrative decisions than in algorithmic predictions; and both transparency and discretion are more likely in single-organization automation than in information infrastructures.
2. Control problems in human-algorithm interactions: bounded rationality, satisficing behaviour, automation bias and frontline coping mechanisms play a crucial role in the way humans make use of the oversight and override options built into algorithms. Algorithms work upon behavioural mechanisms of decision-making and, thereby, limit the ability of humans to control the processes and outcomes of algorithms.

These findings reflect the relevance of insights from science and technology studies regarding the politics of technology (Winner, 1980; Latour, 1987). The agency of algorithms in public administration and public policy refers to the power of algorithms to influence human behavior and decision-making beyond their computer code and immediate calculations. On the one hand, limitations for people's ability to control algorithms may be designed into algorithms and their organizational context. On the other hand, control problems may be a consequence of human-algorithm interaction and behavioural mechanisms that make humans follow the script for action provided by an algorithm. The problem is not only what algorithms do to people, but also what people do with algorithms.

Even though the literature review presented here does not pretend to be exhaustive, it is clear that much empirical work remains to be done to understand the agency of algorithms, the profound changes that algorithms trigger in public administration and government, and the way citizens are affected by these changes. Based on the literature review, three areas of special interest for future research are suggested here:

1. Predictive algorithms: automation has moved beyond process automation of administrative 'production lines' and is, through statistical big data analysis and machine learning, increasingly used for preventative purposes in areas ranging from tax fraud to criminal justice and from predictive policing to child protection.
2. Automated network decisions: automation has moved beyond the single organizational scope and is increasingly used at the level of information architecture, thereby separating data owners and data user organizations and creating new dynamics in administrative decision-making.

3. Frontline automation: automation has moved beyond routinized administrative tasks and is in-increasingly used to support, manage and inform non-routine tasks of frontline professionals in, for instance, welfare work, policing and criminal justice.

In terms of practical relevance, a key question is how to design meaningful oversight and override into algorithmic applications. The conclusions presented here suggest that, besides the organizational challenges that information architectures pose, it is crucial to take into account behavioural mechanisms. A detailed analysis of this issue goes beyond the scope of this article, but one crucial element can be highlighted. In human-machine systems studies, it is well known that, as technology advances, human operators need to transform into supervisors (Greenlee et al., 2018; Strauch, 2018). The financial trader monitors the automatized execution of his strategies and the pilot monitors the proper functioning of an airplane (Zerilli et al., 2019). In public service delivery and frontline work, this transformation has hardly occurred. Without a basic understanding of the algorithms that screen- and street-level bureaucrats have to work with, it is difficult to imagine how they can properly use their discretion and critically assess algorithmic procedures and outcomes. As a means to mitigate the control problems identified above, professionals should have sufficient training to supervise the algorithms they are working with. In other words: "the more advanced a control system is, so the more crucial may be the contribution of the human operator" (Bainbridge, 1983: 775).

References

AI Now Institute (2018). *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems*. https://ainow institute.org/litigatingalgorithms.pdf (accessed March 20, 2020).

Akrich, M., & Latour, B. (1992). A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies. In: Bijker, W., & Law, J. (eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: The MIT Press, pp. 259-264.

Aneesh, A. (2006). *Virtual Migration: The Programming of Globalization*. Durham: Duke University Press.

Aradau, C., & Blanke, T. (2017). Governing others: Anomaly and the algorithmic subject of security. *European Journal of International Security*, *3*(1), 1-21.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775-779.

Banks, V.A., Plant, K.L., & Stanton, N.A. (2018). Driver error or designer error: Using the perceptual cycle model to explore the circumstances surrounding the fatal tesla crash on 7th May 2016. *Safety Science*, *108*, 278-285.

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*, *28*(7), 806-822.

Berendt, B., & Preibusch, S. (2017). Toward accountable discrimination-aware data mining: The importance of keeping the human in the loop-and under the looking glass. *Big Data*, *5*(2), 135-152.

Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & Technology*, *31*(4), 543-556.

Binns, R. (2019). Human Judgement in Algorithmic Loops: Individual Justice and Automated Decision-Making. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3452030 (accessed February 12, 2020).

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

Bowker, G.C., & Star, S.L. (2000). *Sorting Things Out*, Cambridge, MA: The MIT Press.

Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. In: *INTERACT Conference*, pp. 17-24.

Bullock, J.B. (2019). Artificial intelligence, discretion, and bureaucracy. *American Review of Public Administration*, *49*(7), 751-761.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1-12.

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

Citron, D., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, *89*(1), 1-33.

Cordella, A., & Tempini, N. (2015). E-government and organizational change: Reappraising the role of ICT and bureaucracy in public service delivery. *Government Information Quarterly*, *32*(3), 279-286.

Couldry, N., & Mejias, U.A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, *20*(4), 336-349.

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, *29*(3), 245-268.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-40.

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead. Long live digital-era governance. *Journal of Public Administration Research and Theory*, *16*(3), 467-494.

Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law & Technology Review*, *18*, 18-84.

Ekbia, H., Mattiolo, M., Kouper, I., Arave, G., Ghazinejad, A., Suri, R., Tsou, A., Weingart, S., & Sugimot, C.R. (2015). Big data, bigger dilemmas: A critical review. *Advances in Information Science*, *68*(8), 1523-1545.

Elish, M.C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Science, Technology, and Society*, *5*, 40-60.

Engin, Z., & Treleaven, P. (2019). Algorithmic government: Automating public services and supporting civil servants in using data science technologies. *The Computer Journal*, *62*(3), 448-460.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York, NY: St. Martin's Press.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fosch-Villaronga, E. (2019). Responsibility in Robot and AI environments. Working Paper eLaw 2019/02.

Fountain, J.E. (2004). *Building the virtual state: Information technology and institutional change*. Washington D.C.: Brookings Institution Press.

Gillingham, P. (2016). Predictive risk modelling to prevent child maltreatment and other adverse outcomes for service users: Inside the 'black box' of machine learning. *British Journal of Social Work*, *46*(4), 1044-1058.

Goel, S., Rao, J.M., & Shroff, R. (2016). Personalized risk assessments in the criminal justice system. *American Economic Review: Papers & Proceedings*, *106*(5), 119-123.

Goldfinch, S. (2007). Pessimism, computer failure, and information systems development in the public sector. *Public Administration Review*, *67*(5), 917-929.

Gray, C.M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A.L. (2018). The Dark (Patterns) Side of UX Design. In: *CHI Conference on Human Factors in Computing Systems*, pp. 1-14.

Greenlee, E.T., DeLucia, P.R., & Newton, D.C. (2018). Driver vigilance in automated vehicles: Hazard detection failures are a matter of time. *Human Factors*, *60*(4), 465-476.

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

Hannah-Moffat, K. (2013). Actuarial sentencing: An "unsettled" proposition. *Justice Quarterly*, *30*(2), 270-296.

Hannah-Moffat, K. (2016). A conceptual kaleidoscope: Contemplating 'dynamic structural risk' and an uncoupling of risk from need. *Psychology, Crime & Law*, *22*(1–2), 33-46.

Hannah-Moffat, K. (2018). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. Theoretical Criminology. doi: 10.1177/1362480618763582.

Hannah-Moffat, K., Maurutto, P., & Turnbull, S. (2009). Negotiated risk: Actuarial illusions and discretion in probation. *Canadian Journal of Law & Society*, *24*(3), 391-409.

Harcourt, B.E. (2007). *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. Chicago: Chicago University Press.

Houser, K., & Sanders, D. (2017). The use of big data analytics by the IRS: Efficient solutions or the end of privacy as we know it. *Vanderbilt Journal of Entertainment and Technology Law*, *19*(4), 817-872.

Howard, J.J., Rabbitt, L.R., & Sirotin, Y.B. (2020). Human-algorithm teaming in face recognition: How algorithm outcomes cognitively bias human decision-making. *PLoS ONE*, *15*(8): e0237855. doi: 10.1371/journal.pone.0237855.

Høybye-Mortensen, M. (2015). Decision-making tools and their influence on caseworkers' room for discretion. *The British Journal of Social Work*, *45*(2), 600-615.

Introna, L., & Wood, D. (2004). Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society*, *2*(2/3), 177-198.

Janssen, M., & Van den Hoven, J. (2015). Big and open linked data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, *32*(4), 363-368.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*(2), 313-327.

Keddell, E. (2019). Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice. *Social Sciences*, *8*(10), 281-303.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. London: Sage.

Kroll, J.A., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G., & Yu, H. (2016). Accountable algorithms. *University of Pennsylvania Law Review*, *165*, 633-705.

Landsbergen, D. (2004). Screen level bureaucracy: Databases as public records. *Government Information Quarterly*, *21*(1), 24-50.

La Porte, T.M., Demchak, C.C., & De Jong, M. (2002). Democracy and bureaucracy in the age of the web. *Administration & Society*, *34*(4), 411-446.

Le Sueur, A. (2016). Robot Government: Automated Decision-making and its Implications for Parliament. In: Horne, A., & Le Sueur, A. (eds.), *Parliament: Legislation and Accountability*. Oxford: Hart Publishing.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Milton Keynes: Open University Press.

Lipsky, M. (1981). *Street-Level Bureaucracy. Dilemmas of the Individual in Public Services*. New York: Russel Sage Foundation.

Margetts, H. (1999). *Information Technology in Government: Britain and America*, London: Routledge.

Mashaw, J. (1983). *Bureaucratic Justice: Managing Social Security Disability Claims*. New Haven: Yale University Press.

Matzner, T. (2017). Opening black boxes is not enough-data-based surveillance in discipline and punish and today. *Foucault Studies*, *23*, 27-45.

Maynard-Moody, S., & Musheno, M. (2003). *Cops, teachers, counselors: Stories from the front lines of public service*. Ann Arbor: The University of Michigan Press.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1-21.

Monahan J., & Skeem, J. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, *12*, 489-513.

Moon, M.J., & Welch, E.W. (2005). Same bed, different dreams? A comparative analysis of citizen and bureaucrat perspectives on e-government. *Review of Public Personnel Administration*, *25*(3), 243-264.

Moynihan, D.P., & Lavertu, S. (2012). Cognitive biases in governing: Technology preferences in election administration. *Public Administration Review*, *72*, 68-77.

Norris, P. (2001). *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. New York: Cambridge University Press.

O'Keeffe, M., Traeger, A.C., Hoffmann, T., Ferreira, G.E., Soon, J., & Maher, C. (2019). Can nudge-interventions address health service overuse and underuse? Protocol for a systematic review. *BMJ Open*, *9*(6), e029540.

Ostrom, V. (1996). Faustian bargains. *Constitutional Political Economy*, *7*, 303-308.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Boston: Harvard University Press.

Pazouki, K., Forbes, N., Norman, R.A., & Woodward, M.D. (2018). Investigation on the impact of human-automation interaction in maritime operations. *Ocean Engineering*, *153*, 297-304.

Peeters, R., & Schuilenburg, M. (2018). Machine justice: Governing security through the bureaucracy of algorithms. *Information Polity*, *23*(3), 267-280.

Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture – the case of the dutch civil registry's master data management. *Government Information Quarterly*, *35*(2), 175-183.

Ponce, J. (2005). Good administration and administrative procedures. *Indiana Journal of Global Legal Studies*, *12*(2), 551-588.

Postman, N. (1985). *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. New York: Penguin.

Ranchordás, S. (2019). Nudging citizens through technology in smart cities. International Review of Law, Computers & Technology. doi: 10.1080/13600869.2019.1590928.

Scholta, H., Mertens, W., Kowalkiewicz, M., & Becker, J. (2019). From one-stop shop to no-stop shop: An e-government stage model. *Government Information Quarterly*, *36*(1), 11-26.

Schuilenburg, M., & Peeters, R. (forthcoming) (eds.). *The Algorithmic Society: Power, Knowledge and Technology in the Age of Algorithms*. London: Routledge.

Silver, E. (2000). Actuarial risk assessment: Reflections on an emerging social-scientific tool. *Critical Criminology*, *9*(1–2), 123-143.

Simmons, R. (2018). Big data, machine judges, and the legitimacy of the criminal justice system. *U.C. Davis Law Review*, *52*(2), 1067-1118.

Simon, H.A. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. New York: The Free Press.

Smith, G.J.D., Bennett Moses, L., & Chan, J. (2017). The challenges of doing criminology in the big data era: Towards a digital and data-driven approach. *The British Journal of Criminology*, *57*(2), 259-274.

Smith, G.J.D., & O'Malley, P. (2017). Driving politics: Data-driven governance and resistance. *The British Journal of Criminology*, *57*(2), 275-298.

Strauch, B. (2018). Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, *48*(5), 419-433.

Tummers, L.G., Bekkers, V., Vink, E., & Musheno, M. (2015). Coping during public service delivery: A conceptualization and systematic review of the literature. *Journal of Public Administration Research and Theory*, *25*(4), 1099-1126.

Van Eck, M. (2018). *Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming* (dissertation). Tilburg: Tilburg University.

Van Eijk, G. (2017). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. *Punishment & Society*, *19*(4), 463-481.

Van Eijk, G. (forthcoming). Algorithmic reasoning: The production of subjectivity through data. In: Schuilenburg, M., & Peeters, R. (eds.), *The Algorithmic Society: Power, Knowledge and Technology in the Age of Algorithms*. London: Routledge.

Villani, C. (2018). *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*. https://ec.europa.eu/knowledge4policy/publication/meaningful-artificial-intelligence-towards-french-european-strategy_en (accessed March 10, 2020).

Werth, R. (2017). Individualizing risk: Moral judgement, professional knowledge and affect in parole evaluations. *British Journal of Criminology*, *57*(4), 808-827.

Widlak, A., & Peeters, R. (2018). *De digitale kooi*. Den Haag: Boom Bestuurskunde.

Widlak, A., & Peeters, R. (2020). Administrative errors and the burden of correction and consequence: How information technology exacerbates the consequences of bureaucratic mistakes for citizens. *International Journal of Electronic Governance*, *12*(1), 40-56.

Widlak, A., Van Eck, M., & Peeters, R. (forthcoming). Towards Principles of Good Digital Administration: Fairness, Accountability and Proportionality in Automated Decision-Making. In: Schuilenburg, M., & Peeters, R. (eds.), *The Algorithmic Society: Power, Knowledge and Technology in the Age of Algorithms*. London: Routledge.

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121-136.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, *12*(4), 505-523.

Young, M., Bullock, J., & Lecy, J. (2019). Artificial discretion: A framework for understanding the impact of artificial intelligence on public administration and governance. *Perspectives on Public Management and Governance*, *2*(4), 301-313.

Zalnieriute, M., Moses, L.B., & Williams, G. (2019). The rule of law and automation of government decision-making. *The Modern Law Review*, *82*, 425-455.

Zarsky, T.Z. (2011). Governmental data-mining and its alternatives. *Penn State Law Review*, *116*, 285-330.

Zerilli, J., Knott, A., Maclaurin, J., & Cavaghan, C. (2019). Algorithmic decision-making and the control problem. *Minds & Machines*, *29*, 555-578.

Zouridis, S., Van Eck, M., & Bovens, M. (2020). Automated Discretion. In: Evans, T., & Hupe, P. (eds.), *Discretion and the Quest for Controlled Freedom*. London: Palgrave Macmillan.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*, 75-89.

Zuurmond, A. (1994). *De infocratie. Een theoretische en empirische heroriëntatie op Weber's ideaaltype in het informatietijdperk*. Den Haag: Phaedrus.