Towards Principles of Good Digital Administration: Fairness, Accountability and Proportionality in Automated Decision-Making

Arjan Widlak
Marlies van Eck
Rik Peeters

ABSTRACT
Governments are increasingly using algorithms to automate decision-making procedures and generate administrative decisions that previously required human oversight and judgement. However, the black boxing of algorithmic procedures and the automation of street-level discretion lead to concerns about administrative organizations’ ability to balance the need for universal and predictable rules with the need for fairness, proportionality and accountability in individual administrative decisions. In this contribution, we apply classic principles of good administration to automated decision-making and propose additional safeguards to mitigate the risks of automation for citizens. A principle-based approach to the use of algorithms is needed to design in reviewability of procedures and give citizens means to hold administrative organizations accountable for their decisions. However, for principles of good digital administration to have an actual effect, fundamental challenges remain in the institutionalization of these principles and in dealing with the epistemological opacity of algorithms.

1. Introduction

Automated decision-making is gaining ground in public administration and fundamentally changing core bureaucratic procedures and mechanisms. We are already seeing applications in the delivery of social services, in data sharing among government organisations, in tackling welfare fraud, in regulatory practices, in risk assessments, in evaluation of professionals and in determining eligibility for social programmes or taxation (e.g. Bovens & Zouridis, 2002; Berk & Bleich, 2013; Harcourt, 2015; O’Neil, 2016; Smith & O’Malley, 2017; Peeters & Widlak, 2018; Peeters & Schuilenburg, 2018; Van Eck, 2018; Yeung, 2018). There is little doubt that in the near future the automation of decision-
making procedures will expand to more areas and practices. Organisational efficiency is usually the main argument in favour of automated decision-making. Its consequences, however, extend far beyond that – and are, in part, still unforeseen. The implementation of new technology is not a neutral intervention, nor something that merely affects the means through which decisions are made (Verbeek, 2006). In analogy with McLuhan’s (1964) adage that ‘the medium is the message’, the characteristics of a medium or a technology are – perhaps even more so than the content it transmits – key to understanding its social, political and administrative consequences.

Without disregarding potential benefits, we take as a starting point the increasing amount of literature that shows how algorithms and automated decision-making cause problems with fair treatment of citizens in governmental decision-making (Rehavi & Starr, 2014; Grimmelikhuijsen & Meijer, 2014; Smith et al., 2017). Furthermore, we underscore the argument that the use of these new technologies is essentially a tool for further rationalization and standardization (e.g. Pasquale, 2015; Peeters & Schuilenburg, 2018). Instead of applying general rules to individual cases through bureaucratic procedures and case assessment, algorithms are used to automatically generate decisions, determine what cases are handled automatically or not (Zouridis et al., 2020:16) and generate predictions based on statistics that direct attention, provide a default for human decision makers (Hamilton, 2015: 49) or both.

Discretion, regulated by principles of good administration, and the design of administrative procedures used to be the two most important mechanisms to find a balance between the need for universal and predictable rules that allow us to build systems, either organizational or algorithmic, and the need to deviate from those rules guided by principles – such as fairness, proportionality and accountability – that allow us to do justice when the rules fall short. Both these mechanisms are changing because of the automation of street-level discretion (Zouridis et al., 2020) and the black-boxing of algorithmic procedures (Peeters & Widlak, 2018).

A discussion about the legitimacy and fairness of automated decision-making requires an inquiry of two main questions. The empirical question is the comparison of values institutionally safeguarded earlier by classic bureaucratic procedures and now by automated decision-making. This is relevant because the picture of the impact of this technology is far from complete. The change in technology not only changes discretion or procedural transparency, but also shifts power, administrative burdens, bias and more. The normative question is what government should do to
materialize principles of good administration that safeguard these values in this new situation. This may require complementary institutional arrangements.

Rethinking the principles of good administration in automated decision-making is a necessary, but perhaps not sufficient means to ensure legitimacy and fairness. The use of algorithms creates a specific dynamic that we still fail to fully grasp and may well lie beyond control through a reconfiguration of classic Weberian mechanisms. However, contrasting automated decision-making with traditional bureaucratic decision-making helps to understand the values at stake and to define a minimum of safeguards for individual justice.

In this contribution, we will use these two elements as a starting point. We argue that this balance is being threatened by automated decision-making and that, in response, a more principle-based approach of automated decision-making is required. This issue has been raised in various recent studies (Danaher, 2016; Kool et al., 2017; Van Eck, 2018; Widlak & Peeters, 2018; Meijer et al., 2019), but requires further structural elaboration. In the following, we first analyse how automated decision-making threatens to usurp key principles of good administration. Second, we present a preliminary shared set of established principles in advanced democracies that govern administrative decision-making and apply this to automated decision-making. Third, we move from positive law to an inductive analysis of the specific nature of automated decision-making to assess where additional safeguards need to be developed. We conclude this contribution by reflecting on the meaning of algorithms for the way governments provide citizens with access to rights and services.

2. Automated decision-making and good administration

2.1. Automated decision-making in government

Automated decision-making, also known as its near-synonyms of autonomous decision-making, algorithmic decision-making and data-driven decision-making, refers to administrative decisions issued by a government organization to affect the relation between government and citizen. It involves breaking down a decision in a set of ‘if then’ rules and criteria through algorithms (a sequence of reasoning) that make selections from predetermined alternatives (Le Sueur, 2015: 2). By using the word automated, we mean that a computer reaches a conclusion by itself based on data, without human interference (Larus et al., 2018: 2), based on either a pre-programmed
algorithm or on an algorithm that adapts itself through machine learning. The absence of human interference in decision-making can be understood as a continuum (Citron & Pasquale, 2014; Binns, 2016; Danaher, 2016) – ranging from automated decisions that leave virtually no space for human agency (Peeters & Widlak, 2018) to automated decisions that provide a default for human decision-makers, which may be overruled (Hamilton, 2015: 49).

An important distinction should be made between automated individual decisions and automated prediction (EU, 2016). The former refers to decisions that affect individual citizens in their status as either eligible for services or rights or as eligible for enforcement of obligations (such as taxation). The latter is a form of statistical analysis used to identify individuals from a broader group based on specific characteristics that justify the singling out of individuals for further attention. This can be attention in a positive way – the provision of better public services based on profiles – however, profiling is mostly used for enforcement purposes, such as tackling fraud or sorting out citizens that are more likely to show deviant behaviour. The ‘automated’ part is the risk analysis, which is commonly followed up by civil servants (Houser & Anders, 2017: 13).¹

Another key distinction is whether automated decision-making is confined to a single organization or affects the operations and decisions of multiple organizations. The latter implies that data used to make automated decisions might come from a different organization than the one that makes the actual administrative decision that affects a citizen in his rights or obligations, or that decisions made by one organization also impact the administrative decisions other organizations make. The automatic exchange of data leads to chain-decisions (Van Eck, 2018) or automated network decisions. We speak of a ‘chain’ if several hierarchically independent organizations cooperate in a sequential process towards a collective result (Grijpink, 1997; Borst, 2019: 24), such as the police and the prosecution in criminal law. In such cases there usually is a shared legal framework and (data-)definitions are harmonized.

However, the analogy with a supply chain is lost when there is no sequential process, no collective problem or no harmonized definitions. In such cases it is better to speak of automated network decisions. A good example are the so-called ‘basis registrations’ that the Dutch government uses to digitize and centralise vital data on citizens, businesses, geographical locations, buildings, vehicles, and so on. (Peeters & Widlak, 2018). The authentic data contained in these vital registrations as well as changes in that data is automatically shared among a large variety of public

¹ There are, beyond the scope of this contribution, many more government tasks that involve automation, such as digital monitoring of water management systems or managing traffic flows.
and semi-public organizations to allow for the execution of their primary processes. This improves efficiency and reduces double registrations and obsoleteness of the data object. However, it also reduces technical transparency for citizens and user organizations, because the source of data as well as the procedures followed in the use of data often remain unclear (Van Eck, 2018).

2.2. Principles of good administration

Automated decision-making, in its various forms, profoundly alters the way government organizations make decisions regarding the access to rights and services of citizens. These decisions – administrative decisions – are subjected to norms and principles to ensure the protection of citizens’ rights and weigh values in concrete situations and decisions. Given the ‘absence’ of the legislator in the day-to-day execution of the law by administrative bodies, guidelines for administrative decision-making have developed over centuries (Ostrom V., 1996: 1) and are, especially in civil law countries, commonly incorporated in formal administrative law to govern 1) the legality and procedures of administrative decisions and 2) the protection of citizens against breaches of those legal norms (Van der Heijden, 2001). Jurisprudence and Ombudsmen are additional sources of norm setting. In some cases, such as in Dutch administrative law, the EU’s Code of Good Administrative Behaviour, and article 41 of the EU Charter of Fundamental Rights (‘the right to good administration’), these norms have been codified. Their purpose is to guarantee administrative justice and protect citizens against unreasonable or unmotivated administrative decisions – thereby going further than merely guaranteeing the legality of decisions and procedures.²

In the Western legal tradition, where the citizen enjoys legal protection against state intervention and any such intervention needs to be legally justified, principles of good administration are a cornerstone of the rule of law (Mashaw, 2007: 99; Van Hout, 2019) – despite clear differences between national legal traditions. It can be safely argued that administrative decisions made by computers instead of human agents should be covered by the same norms or guidelines as traditional man-made decisions. The Danish case provides a good illustration. Here,

² Furthermore, principles for administrative decisions can also be seen as an instrument to facilitate application of legal frameworks to changing social circumstances (Van der Heijden, 2001). Principles differ from rules in the sense that rules, applied to facts, imply an imperative for action, whereas principles have an open character and allow for interpretation in concrete situations (Dworkin, 1985). Thereby, principles – even if they are codified in positive law – allow for law to remain open for change and adaptation to new circumstances or to specific situations that cannot be captured in general rules.
the Principle of Administrative Law by Design was developed to ensure that legislation and
unwritten principles of administrative law are to be embedded in the technology, practices and
organizational structure (Motzfeld & Naesborg-Andersen, 2018: 139-140). The underlying idea is to
prevent the violation of administrative law due to impetuous design of technologies. This is also
reflected in the European Union’s General Data Protection Regulation, which seeks the protection
of natural persons with regard to the processing of personal data (EU, 2016).

Principles of good administration regulate the use of discretion in individual administrative
decisions as well as the design of fair and transparent administrative procedures. These are the two
prime mechanisms through which rules of law balance universal and predictable rules with the
specifics of individual cases. Crucially, these are also the two mechanisms primarily affected by the
introduction of automated decision-making. First, when decisions or predictions are processed
through computer-programmed algorithms, this leads to a ‘hiddenness concern’ (Danaher, 2016):
it is unknown which data is collected, data is often used without consent and the procedures
through which data is analysed are not transparent (Grimmelikhuijsen & Meijer, 2014; Hannah-
Moffat, 2018), thereby hiding ethical dimensions of analysis and classification from sight (Kallinikos,
2005; Cordella & Tempini, 2015; Peeters & Widlak, 2018). Second, automated decision-making
reduces discretionary space at street-level (Bovens & Zouridis, 2002; Landsbergen, 2004; Zouridis et
al., 2020) – anywhere on a continuum from creating a default or advice for human decision-
makers to full automation in which algorithms decide without human oversight or override (Citron &
Pasquale, 2014; Peeters & Schuilenburg, 2018).

2.3. Towards principles for digital administrative decisions
In the following, we apply existing principles of good administration to automated decision-making
and analyse whether additional safeguards need to be formulated. First, we distil three principles
from an analysis of positive law. Even though every country’s specific legal tradition has given rise
to a wide variety of principles for good administration, we argue that principles of due process,
accountability and proportionality are widely shared. These three principles are applied to
automated decision-making to demonstrate their relevance and usefulness. An important limitation
of this strategy, however, is that it does not take into account the specific logic of automated
decision-making and the consequences it can have for citizens. According to Van der Heijden (2001),
drawing on the works of Dworkin and Habermas, principles of good administration set requirements
for “the action situation”. An action situation is an abstraction of concrete situations where citizens’
access to rights and services is at stake and “where participants [...] interact [...], solve problems, dominate one another or fight” (Ostrom E., 2005: 14). Therefore, we use a second, inductive strategy to analyse how automated decision-making changes this action situation. Through three short case studies, we demonstrate possible undesirable consequences of automated decision-making for a citizen’s legal position. This allows us to identify additional norms to ensure good automated administrative decision-making.

3. Principles in positive law

In the following, three internationally well-established principles of good administration are applied to automated decision-making and automated predictions: 1) due process (which focuses on the procedure of a decision), 2) accountability (which focuses on the allocation of responsibility) and 3) proportionality (which focuses on the material fairness of a decision). The selection of these principles is an attempt to capture a minimum level of shared norms in the variety of national legal traditions in advanced democracies.

3.1. Principle of due process

Due process obliges administrative organizations to demonstrate and make explicit what steps have been taken during the entire decision-making procedure. This principle is, therefore, closely related to the issue of algorithmic transparency (Grimmelikhuijsen & Meijer, 2014; Mittelstadt et al., 2016), which is concerned with the throughput legitimacy of automated decision-making (Schmidt, 2013). Though due process clauses have a different emphasis and impact in different countries, we can say that one essential part of due process is the ability for the citizen to defend himself against an adverse decision by government (Ponce, 2005: 582). This is, for instance, codified as part of the right to good administration in the EU Charter of Fundamental Rights (article 41, sub 1 and 2), which states that every person has to right to have his or her affairs handled impartially, fairly and within a reasonable time, including the right of every person to be heard before any adverse individual measure is taken (EU, 2012). Applied to automated decision-making, due process means that the black box of algorithms should be opened up to give citizens insights in the criteria used and steps taken to reach a decision.
Guaranteeing due process is also relevant for automated prediction (Citron & Pasquale, 2014). Both individual automated decisions and automated prediction have the characteristic that a person is described by data in databases. However, where individual decisions work with business rules programmed by a team of human programmers combined with facts represented by individual data, automated predictions rely on probability derived from statistics, rating and classification (ibid.). Governments mine personal information to make predictions about individuals on questions like: ‘which individuals will be more likely to own crypto currencies and might not report these – either by mistake or intent – to the Tax Authority?’ Essentially different from individual data processing is the addition of an inference based on probability. It is a score that says if a citizen looks like a terrorist, fraud or hooligan. Introduction of the principle of due process implies at least that the quality of the process can be audited (what steps were taken and were they legitimate?) in all the different stages of the scoring process.

3.2. Principle of accountability

Every organization that makes administrative decisions is responsible for them, ‘owns’ them and can be held accountable for them (Bovens, 2010). Public organizations cannot simply hide behind general rules and claim they are ‘merely implementing them’, nor divert responsibility for administrative decisions to individual civil servants. Automated decision-making can raise the same instinctive administrative reaction: ‘the computer says no’ (Van Eck, 2018) – signalling a lack of willingness or capability to take responsibility for individual administrative decisions. Availability of data or lack thereof does not remove the responsibility to verify if this data reflects the actual relevant facts. And technological complexity or poor information management should not be a reason to remove administrative accountability (Fosch-Villaronga, 2019). Some authors consider accountability the twin-sister of transparency. Even if full transparency is not possible, the accountability of the decision-making process as a whole still needs to be ensured (Zalnieriute et al., 2019). However, as Bovens (2010) warns, the principle of accountability may in its institutional implementation lead to proceduralism that hampers reflexivity, efficiency and effectiveness of administrative organizations and lead to goal displacement.

Every administrative organization that uses automated decision-making and automated predictions can be held accountable for the system, its flaws (glitches), the use of the system (such as function creep), and the outcome in individual cases. This includes issues that may be organized beyond the direct control of the administration, such as problems with the code of an information
system that is outsourced. No matter what technology does or how the contract works between the company that sold the technology and the government, the administrative organization is accountable for using the technology and accountable for any actions aimed at individuals. This includes automated network decisions, where organizations make decisions based on data gathered and administrated by other organizations. The organization that makes the decision is accountable for the data it uses – for its correctness and for the consequences of the decisions made based on that data. In other words, the responsibility for automated decisions cannot be transferred to other organizations or to the information system that generated the decision.

The principle of accountability affects both data and procedures. When data is no longer provided by the citizen or to the citizen on paper, the administration must be able to provide the source, the date and content of the data on which a decision was based (Widlak & Peeters, 2018: 111-117). When procedures are no longer (practically) accessible, the administration must be able to organize a procedural review in its professional community and professional discourse. Consequently, correction of erroneous decisions must be a possibility. Data and software cannot be interwoven in a way that makes this a practical impossibility (ibid.: 117-120). Keeping ‘humans in the loop’ does not mean that individual civil servants should constantly verify automated decisions, but the principle of accountability does imply a right to a ‘human eye’ in complex cases, that allows for comparison of data with the facts in real life and a meaningful intervention in individual cases (see also article 22 of the EU’s General Data Protection Regulation; EU, 2016).

3.3. Principle of proportionality
The exclusion of human agency in decision-making begs the question if an organization can still guarantee that the decisions it makes lead to fair results and do not affect citizens excessively or unnecessarily (cf. Ranchordás & De Waard, 2016). The principle of proportionality can be aimed at the legislator when it wants to intervene in human rights (such as Article 8 of the European Convention of Human Rights and Fundamental Freedoms), but it can also be used to guide the behaviour of civil servants (see article 6, section 1 of the European Code of Good Administrative Behaviour; EU, 2002). In automated decision-making, proportionality needs to be guaranteed at both levels as well. But given the fact that a single set of data may now be used for a myriad of administrative decisions, the legislator should assess proportionality at the level of information architecture as well: are all governmental interventions combined aimed at one citizen still in balance? Or should strategies be developed to mitigate possible disproportional consequences?
At an individual level, proportionality first and foremost requires the possibility to overrule an automated decision when circumstances are unforeseen or errors have been made. Generally, the use of automated decision-making has shifted human judgement from the core of the decision-making process to the phase of objection and complaints, thereby shifting the responsibility to observe the need for judgement and to monitor for errors from the administration to the citizen. This, however, is not a necessity. Just like automated predictions can be used to search for fraud, they can also be used to find judgement cases for example.

The issue of proportionality is especially relevant in cases of automated network decisions. In the execution of tasks, administrations are using data that are provided by other agencies for their own primary decision-making processes. For instance, the aforementioned ‘basis registrations’ require Dutch governmental agencies to use data already available in designated governmental databases. Other countries are also working on implementing the ‘single registration, multiple use’ principle as an essential part of their e-government strategies (e.g. Digital Government Factsheet Norway, 2019: 7; Government of Flanders, 2015: 4, 17-18). A downside of these strategies is that user organizations often have little interest in carefully monitoring the quality of the data they receive. Furthermore, data are not always interoperable – the definitions in the different applicable laws are not the same, but their databases are used to compile the decisions nevertheless (Van Eck, 2018: 447). Another issue in this respect is that the consequences of a change in vital records for an individual citizen cannot be fully foreseen (Peeters & Widlak, 2018), which complicates the assessment of the proportionality of such an administrative decision.

4. Good digital administration in changing action situations

In this section, we present three short cases of citizens that have been negatively affected by automated administrative decisions. More specifically, all cases involve automated network decisions, in which algorithms govern the exchange of data from vital public records and, thereby, impact the capability of organizations to guarantee fairness, accountability and proportionality in their individual administrative decisions. The objective of this inductive approach is to assess the need for additional or more specific norms to ensure good digital administration. All cases involve an administrative error, because errors made by the administration are the most undisputed examples of specific circumstances that justify use of discretion and deviation from the rules. The
first two cases are based on original data gathered through document analysis and interviews in 2018 and 2019 in the Netherlands. The third case (see also Widlak & Peeters, 2020) has been studied since 2014.³

4.1. Buy one house, pay property tax for two

When Simone bought a house in Amsterdam, she soon received a property tax assessment. Strangely enough, she also received a second assessment for a house a bit further down the street. She files a formal complaint, which is only handled by the municipality fourteen months later – two months after the legally established maximum period. According to the municipality, research by the municipal department of vital record registries has demonstrated that she is registered as the owner of the second house and, therefore, is required to pay the corresponding property tax. In order to prove the municipal data is incorrect, Simone obtains an extract from the land registry office which shows that the owner of the second house is an investment fund. Records also show that the property has exchanged hands multiple times during the past few years. Eventually, proof is retrieved of the ownership at the time the municipality sent Simone the tax assessment. She sends the documents to the municipal ombudsman and not long after she receives notice that the mistake will be corrected.

Simone’s case is resolved in her favour, but not after a lengthy and complicated procedure in which the burden of proof of not owning a house was placed on her. Analysis of this case shows that the problems can be traced back to the reproduction of an administrative error through automated network decisions. Amsterdam’s municipal property tax is determined by linking three types of data: the land registry’s data of ownership is linked to data on addresses and buildings, which linked to the residents registry. This data is updated frequently and stored in a municipal data pool. In Simone’s case, a human error was made in this process and ownership of a property was incorrectly assigned to her. When Simone filed her complaint, municipal employees reviewed their own data instead of looking at the original data sources or at the evidence presented by Simone. In the land registry, they could have seen that Simone was not the owner of the second house. However, employees trusted the municipal data to be correct and proceeded accordingly. Algorithms provided them with a default which they did not question.

³ A more detailed methodological justification is available upon request and available in Widlak & Peeters, 2020, for the third case.
4.2. Paying for unreceived care

Piet is the legal administrator for his father, who is hospitalized and receives geriatric care rehabilitation. Months before termination of the treatment, it is clear that Piet’s father will require permanent care. The hospital suggests him to file a request for long term care at the Care Assessment Centre (‘CIZ’). This is publicly funded, but also involves a personal contribution. A month after Piet’s request, he receives an indication of eligibility pre-dated a month before the date of his request. Another month later, the actual long-term care starts. A different administrative body, the Central Administration Office (‘CAK’) determines Piet’s monthly personal contribution. Basing itself on data obtained from the Care Assessment Centre, the amount is set on 850 euro – starting three months before the start of the actual treatment. This would mean that Piet has to pay for three months in which no actual care was received. After an unsuccessful complaint, Piet starts a law suit against the Central Administration Office claiming that he should not pay for the three months identified incorrectly as part of the long-term care.

The judge rules that the Central Administration Office may trust the information it automatically receives from the Care Assessment Centre regarding the effective starting date of a personal contribution (Rechtbank Midden-Nederland, 2018). The reasoning is that the Central Administration Office cannot be made responsible for verifying if the assessment centre’s decisions are correct and if the care provided by hospitals and other institutions is in accordance with the information provided. The Central Administration Office’s responsibility is merely to provide the funding for the institutions providing the long-term care and to collect the patient’s personal contribution. While this makes sense form an organizational perspective, it also means that an error made by the assessment centre or care institution is, therefore, reproduced without any form of control and that the burden of correcting it falls on the citizen – if such a correction is practically possible at all. Automated network decisions facilitate the spread of errors and set a default which is not questioned by user organizations.

4.3. A stolen car

On April 30, 1998, Saskia’s car is stolen. The same day she reports this to the police. On August 20, Saskia receives a letter to remind her to have her car tested. Initially, she thinks her police report has not been processed yet. However, soon after that she starts receiving tax forms from the Dutch tax authority. A complaint by Saskia leads to nothing as the tax authority says that data from the vehicle registration authority confirms she is the owner of a car. The vehicle registration authority
claims its records to be correct despite Saskia’s police report of her stolen car. Therefore, she remains liable for motor vehicle tax and vehicle safety tests. Saskia is a single mom and over the years her financial problems accumulate and she has trouble paying her taxes – which she does to avoid legal problems, even though she is convinced she should not have to pay for a car she does not own anymore. When she loses her job, she is no longer able to pay and the judicial collection agency comes into action to collect her debts. This agency also claims the fines are justified based on the data provided by the vehicle registration authority. In 2011, Saskia finally succeeds to strike the title of ownership from the vehicle registration. From now on, no new taxes and fines are added. However, this does not automatically lead to a nullification of all outstanding bills, nor to a refund of all unjustly paid fines and taxes.

Desperately, Saskia sends a handwritten letter to the mayor of Rotterdam in March 2014. This triggers an unofficial investigation, which shows that the vehicle registration authority received an ‘end date of theft’ by the police the day after Saskia’s car was stolen in 1998. Subsequently, the authority reinstated Saskia as owner of a vehicle. After some pressure by the mayor’s office, the police look into the case as well. It turns out that the car was found the day after Saskia had reported it as stolen – the police had simply failed to inform her about this. On September 4, 2014, the police offer a letter of apology to Saskia. Based on that letter, the tax authority is willing to nullify all outstanding taxes. However, the tax authority is unable to look back in their records more than five years nor able to provide a record of road tax settled with other taxes. The vehicle registration authority will not retroactively change Saskia’s registration because this “would severely harm the integrity of the registry”. And the judicial collection agency informs Saskia that she is no longer in their system and, moreover, that they cannot reimburse the fines she paid because “we already sent the money to The Hague years ago”. As a result, Saskia still has debts. Automated decision-making not only causes a simple error to spread throughout an entire system, it also proves highly complex in practice to identify the source of an error and to make organizations assume responsibility for the consequences of an error.

4.4. Analysis

In all cases, the combination of an administrative error and automated network decisions triggers enormous problems for individual citizens. Specifically for network decisions – in which data is

4 ‘The Hague’ refers to the seat of Dutch government, where the Ministry of Justice is located.
automatically shared among organizations – the previously described three principles for good digital administration seem to fall short of providing adequate protection for citizens. The main issue here is that an organization makes administrative decisions based on information coming from other organizations. An organization may formally ‘own’ a decision and bear responsibility for it, but in practice this responsibility is evaded by two mechanisms:

1. Design problem: bureaucratic barriers often impede an organization in both formal and practical terms from verifying the source of the information upon which it depends to make its own administrative decisions. This is perhaps best exemplified by the second case, in which a judge rules that an administrative organization may trust the information it receives from another organization. Crucial elements of a decision-making procedure are obscured for both the ‘owner’ of the data and the ‘owner’ of the administrative decision: the former cannot assess or control the applications and consequences of the data it collects and distributes, whereas the latter cannot verify the data upon which it makes a decision that impacts the life of a citizen. Algorithms govern these processes and blindside all parties involved.

2. Discretion problem: even if algorithms allow for human oversight and override, they may create a behavioural default for civil servants to trust the presumed objectivity of data and hide behind the complexity of information architecture instead of using their discretion for individual case assessment. As the first case shows, civil servants are inclined to trust municipal data over the arguments presented by an individual citizen. ‘Hard facts’ are preferred over ‘real life’ evidence – a strong argument is needed for not following what automated processes present as the truth. Moreover, in its current implementation it often requires and enormous practical and analytical effort to uncover all data sources, links and algorithms behind an automated decision. And as the third case shows, data and the information architecture that produces and distributes are not designed to allow for correction and especially not for retroactive correction. Hence, they are treated as sacrosanct institutions not to be meddled with at the price of ‘loss of integrity’ or as immovable objects that defy any attempt of human intervention.

5. Complementary principles for automated network decisions
The logic of automated network decisions complicates compliance with, above all, the principle of accountability. This principle demands organizations to take full responsibility for the decisions they make – the way an administrative system or information architecture is designed cannot be an excuse to avoid accountability. Yet, this is exactly the reality of many automated network decisions: data collected by other organizations is either inaccessible for citizens, by default presumed correct or cannot be corrected. This, by definition, also affects the two other identified principals for good digital administration. Due process cannot be guaranteed because data collection and data sharing are not made fully transparent for neither the decision-making organization nor the affected citizen. Of particular concern is the difficulty citizens face when trying to identify where and by whom an error has been made. Furthermore, proportionality is at stake for two reasons: 1) the organization that collects and shares data cannot guarantee that decisions made by other organizations will be reasonable and 2) the organization that makes a decision based on erroneous data shifts the burden of proof for detecting errors and correcting their consequences for multiple organizations to the citizen.

In light of our analysis, we argue that automated decision-making places a considerable strain on the ability of an administrative organization to ensure accountability, due process and proportionality in administrative decisions. Therefore, we suggest the introduction of additional elements to the previously identified principles in order to mitigate the specific risks that automated network decisions pose to good digital administration:

1. **Procedural completeness** to complement the principle of due process. The citizens in the three cases presented above face an information problem. The inability to identify the source of administrative errors can be traced back to a lack of insight in what data is used for decision-making, who owns this data and whether this data is complete, contextually relevant, correct and up-to-date. In traditional decision-making, the organization that owned the decision also owned the data. The two have been separated in automated network decisions. Completeness in the process of decision-making means that an administrative organization must be able to provide a citizen with information about both elements of the procedure. This goes beyond traditional calls for algorithmic transparency, because these do not address the division of labour between data collection and sharing on the one hand and individual administrative decisions on the other hand. Both procedures may be transparent individually, but still add up to an opaque system from the perspective of the citizen and the decision-maker.
2. **Factual assessment** to complement the principle of accountability. Administrative organizations avoid responsibility for their decisions because they tend to treat data as unquestionable facts and the information architecture as an untouchable entity. Trusting the algorithms is the default option for decision-making. To prevent organizations from hiding their responsibility, the possibility of a human assessment of administrative decisions must be guaranteed. It is crucial, however, that this assessment is fact-based instead of data-based: instead of reviewing databases, a citizen’s factual situation should be the focus of the assessment.

3. **Central correction** to complement the principle of proportionality. Automated decisions based on erroneous information can lead to disproportional consequences for affected citizens. The burden of proof for administrative errors is placed on the citizen. This is often also the case with traditional paper-based decisions, but an important difference is that correction not only needs to take place in the organization that made the individual administrative decision, but also in the organization that collects and shares the data. Moreover, the consequences of erroneous data increase as this data is shared with other organizations. To prevent citizens from being disproportionally burdened with erroneous or unfair administrative decisions, a central correction of the error must be guaranteed.

6. Conclusion

The classic Weberian bureaucracy in which civil servants individually apply general rules to individual cases is since long a thing of the past. Especially since the 1990s, databases and computers have steadily taken over tasks originally entrusted to paper files and human assessment. The ‘infocracy’ (Zuurmond, 1994) is an important touchstone in the digitalization of administrative decisions. The increasingly widespread use of algorithms marks a new phase in which not only data is digitalized but entire decisions-making procedures are automated. Moreover, automated network decisions combine the ‘algocracy’ (Aneesh, 2006) with the ‘system-level bureaucracy’ (Bovens & Zouridis, 2002). Every technological transformation in administrative decision-making begs the question how this affects citizens in their rights and obligations. The main concerns regarding automation of processes is that it reduces an organization’s capacity to assess individual cases because it automates discretion (Zouridis et al., 2020) or at least generates automated predictions (Lorenz,
2019) as well as an organization’s control over the fairness and transparency of its procedures (Grimmelikhuijsen & Meijer, 2014; Peeters & Widlak, 2018).

Whether we are talking about rights and obligations determined by individual automated decisions, about risks and profiles identified through automated prediction, or about vital public records shared among multiple organizations in automated network decisions – the principles that govern good administrative decision-making are at risk of being usurped by a technology that fundamentally transforms the way administrative organizations make decisions. Due process is at risk because data collection and data sharing are not transparent, accountability is complicated because organizations pass responsibility for the correctness of data on to other organizations, and proportionality is jeopardized because the responsibility for detecting errors and correcting their consequences for multiple organizations is placed on citizens. In response, we have identified three specifications of these principles to mitigate the specific risks that automated decision-making, and especially automated network decisions, pose.

However, formulating principles for good digital administration on paper is something else than safeguarding them in practice. Three challenges need to be tackled. Algorithms pose, first, a design problem because they complicate oversight or ‘reviewability’ (Danaher, 2016) by organizations and affected citizens. Here, a key issue is to design algorithms that keep humans in-the-loop and organize the possibility of oversight and override (Citron & Pasquale, 2014). Second, algorithms trigger a behavioural problem because they set a default for action which – despite keeping humans ‘in the loop’ – is likely to be unquestioned (Peeters & Schuilenburg, 2018). Both problems can only be tackled through a thorough institutionalization of principles of good digital administration. Much work remains to be done to ensure a principle-based design of automated decision-making.

However, there is also a third, epistemological problem in the input, throughput and output of automated decisions. The issue here is not that algorithms are in some way hidden, but that they might be inherently opaque or incomprehensible to human reason (Danaher, 2016). At the level of input, this translates into unknowability of the data used and their origin to reach a decision – especially if ‘big data’ and network decisions are involved. At the level of throughput, algorithms can be relatively simple and ‘interpretable’ or, as is often the case with machine learning algorithms or predictive algorithms, they can be ‘non-interpretable’ and cannot be “reduced to a human language explanation” (Zarsky, 2011: 293). Finally, at the level of output, a decision generated by algorithms can equal a fait accompli for both citizen and human decision-maker, which can only be
questioned ex post. Merely keeping a human in the loop is, therefore, not enough. The question is how this can be done in a meaningful way. How can we organize the analytical tools for humans to assess whether an automated decision was based on correct and unbiased data, was reached according to fair criteria, and how a decision and its consequences can be overturned? And how can human decision-makers identify the exact cases from the massive daily flow of automated decisions where specific circumstances caused unreasonable outcomes? This shows that the challenge goes beyond forcing automated decision-making into the templates of classic ‘analogue’ decision-making. The key challenge lies in developing the mechanisms that are responsive to changing action situations.

References


