Predictive Analytics and Citizens Trust in Government
Scherg, Severin

Master's Thesis: Algorithmic Governance

Predictive Analytics and Citizens Trust in Government

Severin Scherg

S2955466

M.Sc. Public Administration: Economics & Governance

Faculty of Governance and Global Affairs

Leiden University, The Hague

Supervisor: Prof. Dr. ing. A.J. Klievink

Wordcount: 19.460

11th June, 2021

Abstract

The increased amounts of data collected by governments and the public sector enable the usage of novel artificial intelligence applications such as predictive analytics. Citizens provide various information to governments and their administrative arms through a steady interaction. While previous research has focused on how citizens perceive algorithms and more generally the ethical limitations and boundaries of the technology in the governance sphere, this thesis explores how predictive analytics may impact the relationship of trust between citizens and their government. Using primary data collected through an internet-based survey, I present a qualitative as well as a quantitative analysis of different factors such as trust, transparency and a potentially widening power asymmetry between citizens and the government. The survey follows the vignette approach and randomly presents respondents with one of two scenarios that are distinct in their level of transparency to analyse its function as an intermediary for trust.

Loss of trust was the most indicated reason for a negative perception change towards the government in the scenario, followed by a lack of transparency. Respondents from both scenarios indicated this lack of transparency and although the median transparency score differed in the intended direction, the statistical tests were insignificant. Finally, exploring how useful respondents judge algorithms to be, the two largest groups consist of respondents stating that it depends on the usage as well as others that highlight the efficiency gains for the public sector.

Table of Contents

List of Tables

1. Introduction

1.1 Research Background

Accurately predicting an event or a behaviour before it sets in would certainly revolutionize business, government and many other factors present in our daily lives. One example of a successful tool for making predictions is shown in the 2011 film “Moneyball”. Using an empirical analysis tool also known as SABRmetrics, a graduate economist was able to successfully scout undervalued baseball players for his franchise, the Oakland Athletics, leading them to a historical winning streak which revolutionized the sport. Replicating this success story outside of the closed world of baseball and its clear ruleset has proven to be trickier but an ambition which both the private and public sector share.

In the year 2000, three-quarters of the data present in the world was in analogue form. 15 years later, over 99% of data was digital (Mayer-Schönberger, 2015: 788). Although the private sector has learned to embrace the possibilities, which arise through the generation and collection of vast amounts of data, the public sector is still lagging behind (O’Reilly, 2013; Höchtl et al., 2016: 150). The trends which are driving this run on algorithmic and Big Data technology are a combination of increased data collection, the improvement of algorithms as well as enhanced and rising processing power (The Economist 2021). Especially in the field of data collection, the public sector is dominating. Alon Peled explained that “the public sector’s digital data troves are even bigger and growing at a faster rate than those in the private sector” which enables more possibilities for algorithmic and analytical analysis (Peled, 2014 cited in Höchtl, J. et al., 2016: 150).

Key to harnessing new data-driven technologies is the collection and processing of Big Data and the application of artificial intelligence (AI) methods such as machine learning (ML), neural networks and multiple types of analytics. These are set to make a lasting impact on the public sector, although some of these developments have come under scrutiny from the general population, privacy advocates, NGOs and courts. A 2014 report presented by Executive Office of the President of the United States underlined some of the concerns. Janssen & Kuk (2016) conferred the most pressing issues outlined in the report: “Big Data technologies can cause societal harms beyond damages to privacy” plagued by an “opaque decision-making environment”, an “impenetrable set of algorithms”, and notably the dangers of “encoding discrimination in automated decisions" (Executive Office of the President, 2014 as cited in Janssen & Kuk, 2016: 373).

One example of a Big Data analytics application is the System Risk Indicator (SyRI) programme in the Netherlands which was launched in 2013 and stopped indefinitely in 2020 due to multiple legal challenges (De Rechtsspraak, 2020).

The SyRI programme was used to link together different data on citizens and create a model which presented individuals whose profile predicted them to be at risk of committing benefits or tax fraud as well as potential violations against labour laws (Henley & Booth, 2020). In this case, the linkage of data and the algorithmic analysis presented the civil servants with data-driven predictions of "at-risk" individuals. Controversial in the development of algorithms handling such far-reaching decisions and categorizations is their complexity, which opens up the possibility of implanting errors that taint its further development and usage (Waller & Waller 2020: 1). The SyRI application and its underlying algorithm was deemed as opaque and complex, thus having the potential to severely influence the relationship between citizens, civil servants and the state. While there is an organizational need to compensate capacity gaps in assessing individual's eligibility for welfare and monitoring potential fraudulent infringements, using predictive analytics to classify some individuals or neighbourhoods enables the government to set a precedent for warranted mistrust. This may also shift citizens to oppose novel algorithmic-based technologies. An intriguing part of this discussion is the field of tension that arises between efficiency gains and individual privacy infringements, thus detrimentally affecting the relationship between citizens and governments. Sonja Bekker highlights "chilling effects" that untransparent and opaque tools such as SyRI can have on citizens willingness to share their data with public administrations and the impact such applications can have on trust in the government (2021: 299). Moreover, such usage of opaque algorithms and novel technologies poses governments with the challenge of presenting the added value of these systems to citizens and staying within the expected legal and moral boundaries.

The intriguing part of the SyRI programme was the general lack of transparency in the implementation of the programme and the ex-post publication of its usage. The criticisms that were levied against SyRI as well as the general perception of citizens towards predictive analytics in the public sector are much more difficult to entangle after widespread media, political and judicial attention. Therefore, I would like to analyse on a more general level how citizens perceive algorithmic technology such as predictive analytics when used in the public sector, thus interacting and affecting lives on a widespread basis. To further narrow down this perception, I propose an analysis of how predictive analytics may influence the consisting relationship of trust between citizens and their governments. Relationships between people but also between people and institutions are based on a multitude of factors, one of those being trust. Thus, I aim to explore how the relationship changes when a novel technology such as predictive analytics is introduced. Furthermore, usage and "used upon" are also clearly defined, with the government in this case being the user and the citizen the "used upon". Individuals will have varying expectations, understandings and fears towards predictive analytics. I thus assume that these perceptions and understandings will also reflect into the perception that the individual has of the government, more explicitly of the "amount" of trust they place towards the government.

Taking this into account, I will analyse the effect that predictive analytics have on trust and how this could have impediments in the future work of the public sector due to a deterioration of trust. This culminates into the research question:

How does the application of predictive analytics affect the relationship of trust between citizens and the government?

Much of the literature on AI applications and its usage propose mainly theoretical approaches to securing continuous trust in such technologies by focussing on ethical usage and stakeholder consultations. By focussing a further subquestion on transparency, the analysis aims to move away from more theoretical, ethical discussions and propositions towards governmental and organizational actions.

Furthermore, I attempt to uncover if there is general trust or distrust towards predictive analytics or if transparency initiatives and generally a government which values and employs transparency is trusted more with the implementation of Big Data applications. This assumption stems from Stephan Grimmelikhuijsen's (2009) research if transparent government agencies strengthen trust. I choose transparency, due to its growing importance in the realm of "open government", in which organizations are facing "more active demands for disclosure of information" (Oliver 2004: 37 as cited in Park & Blenkinsopp 2011: 256). This growing activity of transparency through information disclosure can lead to multiple positive effects in relationship between citizens and their governments such as correcting poor performance and increasing accountability (Park & Blenkinsopp 2011: 256). Furthermore, and relevant for my research interest and subquestion, Rawlins (2008) states that organizations that are more transparent will also benefit from an increase in trust (as cited in Park & Blenkinsopp 2011: 256). In this context, Park & Blenkinsopp (2011) as well as Welch et al. (2005) describe:

> “Trust in government or public services is typically measured in terms of citizens’ subjective judgements based on their experience, suggesting that citizens’ trust will arise when a government or public service is viewed by citizens as competent, reliable and honest, while also meeting their needs” (2011: 257).

Part of building up this relationship of trust between citizens and governments is thus based on the premise that citizens “sufficiently monitor and control governments’ performance” (Park & Blenkinsopp 2011: 256). Therefore, the second chapter will present the different directions and varieties of transparency as described by Hood & Heald (2006) as well as how transparency may function as a driver or an intermediary of trust, leading me to my second research question:

Does transparency influence citizens attitudes towards a government that uses predictive analytics?

By differentiating if a lack of transparency or the technology per se is the main effect on trust I aim to present citizens’ concerns towards predictive analytics in a nuanced manner and highlight the importance of the different varieties and directions. In this context it will be analysed if the transparency of the institution matters and can potentially “cover” for complex and opaque algorithms and Big Data applications:

1.2 Academic and Societal Relevance

The field of artificial intelligence in the public sector encompasses a broad scope. Examples of scholarly work include research on concepts such as algorithmic governance, the accountability of AI applications and the ethics surrounding its usage to name only a few. In this thesis, the focus lies on predictive analytics and its effect on citizens trust in government. Past works have highlighted the intrinsic opacity as well as the complexity that is inherent to algorithmic applications. Taking this into account, transparency will be regarded as an instrument or intermediary to achieve the outcome of trust thus providing an engaging and multi-faceted research problem.

While the goal in public sector predictive analytics is to uncover meaningful patterns that point toward inferences such as fraud, the broader implications go further than performance enhancement. To successfully make predictions a broad swathe of data is needed to cover a multitude of potentially influential variables. Ambiguous in this approach, especially when placing citizens into a category that states that they are at high-risk of committing fraud such as in the SyRI programme, is that the underlying datasets present the world in the past (Waller & Waller, 2020:

6). Therefore, it is important to assess what performance improvement constitutes in the public sector and how severe potential trade-offs, especially those including novel technologies, could be. Citizens must therefore have an understanding of utilities as well as disutilities that come with these applications. To present and understand these trade-offs, transparency can potentially be a powerful tool to achieve consensus around the implementation and usage of such applications. Therefore, transparency is chosen as an intermediary, as past works in public administration research have highlighted its importance for a functional relationship between citizens and governments. Paving the way for a more data-driven and evidence-based policy making is desirable for governments, but it remains to be seen how citizens judge far-reaching decisions and implications in their own lives that are constructed through Big Data and AI applications and if achieving evidence-based policy making through these means could affect trust. For policy-makers and civil servants, reflection is an important tool which leads them to question that if something is technically possible, should it be done and is its use warranted in the public sector? The European Liberal Forum warns that: "[…] political arguments based on data analytics are likely to favour actors with more skills and resources […]" which could lead to a decline in accountability as decisions framed as evidence-based could alienate citizens trust towards governments (European Liberal Forum, 2019: 30).

Societal values, legal boundaries and ethical considerations are all challenges which the public sector faces on a day-by-day basis and which can lead to difficult trade-offs. To complement the assessment of these difficulties and trade-offs, I will propose the *Society-in-the-loop* concept by Iyad Rahwan (2017), which is conform to a societal wide approach towards the implementation of new Big Data technologies. Taking this into consideration, I present a more general initiative of securing a consensus on the future of algorithmic technology.

Furthermore, a key component of the analysis is the area of conflict arising between potential efficiency gains and privacy intrusion in the field of predictive analytics as well as the resulting effects on the trust between citizens and the state. It is important to take into account that citizens do not have the possibility to opt-out of government and thus cease the continuous provision of data which is of great use for the public sector. This means, that citizens must see the added value that Big Data and analytics deliver to maintain the complex relationship of trust between citizens and the state (Chawda 2018).

1.3 Research Gap

In the past there has been a wide array of studies which investigated the perceptions individuals have on algorithms. Dietvorst et al. (2015) coined the term algorithmic aversion after studying students' reactions to seeing algorithms make errors as opposed to human decision makers. Surveys conducted by Smith (2018) concluded that a majority of the respondents voiced general concerns about using algorithms in decision-making and many deeming it unacceptable (as cited in Araujo et al. 2020: 612). Brown et al. (2019) analysed the perceptions related to algorithmic decision-making in child welfare services and the underlying reasons why such projects have been discontinued in the past, as well as possibilities for the wider community and stakeholders to be included in the development and operation of algorithms (2019: 2). Waggoner et al. (2019) found evidence of a Big Data effect in their study that influenced participants to equate Big Data with quality even if they are not able to fully understand the processes and functions which are then carried out by an algorithm, thus partially opposing Dietvorst et al. (2019: 117). Furthermore, they highlight Kennedy et al. (2018) who proclaimed that the factors which influence individuals trust in automation and algorithms are an important future venue of research (cited in Waggoner et al. 2019: 119).

While most of the referenced studies aim to uncover the individual's perception of algorithms, automated decision making and their influence in specific decisions, the main interest of this thesis is if predictive analytics may influence the relationship of trust between citizens and the state. I focus on this specific algorithm due to its potential to guide decision makers ex-ante, thus possibly leading to action before a behaviour or event has set in. This constitutes an intriguing narrative as to how this influences the relationship between citizens and governments. As I will lay out in the following chapters, aggregating and modelling data to predict a behaviour or an event is accompanied by a specific set of risks that are inherent to the algorithm as well as to the perception of citizens. A past study by Bekker (2021) which conducted an in-depth discussion about the SyRI case further contributed to my interest in predictive analytics. Bekker states that in the SyRI programme, citizens did not know whether they are surveilled and what behaviour is deemed as suspicious, which led me to question how a similar predictive algorithm with a different background would be perceived by citizens and if an intermediary, such as transparency, could potentially change the outcome (2021: 292).

The factors on which this will be focused will be laid out in Chapter 2 and hypotheses will be presented that further narrow down my research interest and stem from the theoretical basis laid out in the chapter. Furthermore, similar to Kennedy et al.'s emphasis on the underlying factors, I propose transparency as an intermediary of trust which may influence citizens perception of the usage of predictive analytics and thus also the relationship of trust between citizens and their government.

The following chapter will present a theoretical framework and introduce the core concepts for this thesis. First, the field of algorithmic governance will be introduced to present the theoretical foundations, the rationale and an assessment behind the push towards algorithmic applications in the public sector. Subsequently, the concepts of Big Data and predictive analytics will be clarified as well as the potential drawbacks. To conclude the theoretical chapter, I will present theories devolving in the concepts trust and transparency. By first highlighting the power asymmetry between citizens and governments, which could potentially be intensified by predictive analytics, I explain why this perception of a power asymmetry may influence citizens in their judgement. Afterwards, transparency and algorithmic systems will be introduced to highlight how transparency as well as opacity are present on the implementational as well as on the operational level. Thereafter, the directions, varieties and outcomes of transparency based on Hood and Heald (2006) are introduced. The outcome of interest in this case is trust, which will also be presented in the subchapter. Following the theoretical chapters and the hypotheses which are derived from the chapter, I will present the research design and methodology. To gather the data, an internet-based survey was distributed which included two distinct, hypothetical scenarios with varying "levels" of transparency. The fourth chapter will put forward the collected data, an in-depth analysis and a discussion of the results. The final chapter will first present a concept and outlook for predictive analytics in the public sector before laying out the strengths and limitations of this thesis. Finally, a summary concludes the findings and reiterates the key take-aways.

2. Theoretical Background

2.1 Introduction

In next chapter I will present the ideas, core concepts and theoretical basis informing the research question of this thesis. The first part of this chapter will lay out the definition and assessment of *Algorithmic Governance* and its relevant subdivisions. Beginning with the concept of *Governance by Algorithms*, I will confer how governments and the public sector aim to integrate algorithmic systems into their governance structures and how governance is to be presumed in the algorithmic age and assessed in the academic literature. Succeeding this, I will introduce *Big Data* and its value as well as implications for algorithmic systems, such as predictive analytics. Thereinafter, the methods, technological features and usage of *Predictive Analytics* will be described.

The part of this chapter titled *Trust and Transparency* connects these topics theoretically to each other as well as to algorithms and more specifically predictive analytics. The first subchapter will introduce the *Power Asymmetry* which algorithmic applications such as predictive analytics enable and potentially magnify in the relationship between citizens and their governments. Thereinafter, the concept of *Algorithmic Aversion* by Dietvorst et al. (2015) will be introduced highlighting a further effect which could stem from public sector usage of predictive analytics. Additionally, theories on transparency will be assembled to point out their importance not only for algorithmic applications but also generally for maintaining and safeguarding citizens trust in government. Among these, I will first present *Transparency and Algorithmic Systems* including a partition between intrinsic and extrinsic opacity. Thereinafter, the *Directions, Varieties and Outcomes of Transparency* will be presented to finalize the theoretical part of the thesis. Drawing mainly on Hood & Heald (2006), this chapter bridges these different aspects and explains how transparency can be used to drive outcomes, in this case trust. Finally, a *Conceptual Framework* which visualizes and brings together the different concepts and theoretical aspects and situates the hypotheses will be presented.

2.2 Algorithmic Governance

2.2.1 Governance by Algorithms

“Government is about the doing and governance is about the abstract structure of what is happening and changing” (Bannister & Connolly 2009: 9 as cited in Höchtl, J. et al. 2016: 147). Governance is a continuous process and an essential part of the way governments react to and implement preferences of citizens and policy programmes. Höchtl et al. describe governance as those decisions made by the government which reflect social expectations through leadership and management (2016: 147).

Algorithms are artificially constructed intelligent applications that have the computing power to manipulate broad swathes of data to achieve an outcome (Sandvig 2014 as cited in Janssen & Kuk 2016: 371). These outcomes qualify social structuring as a mode of governance (Katzenbach & Ulbricht 2019: 2). Past studies on algorithmic systems in governance and administration processes highlight two main focuses: enhancing service delivery as well as informing decision-making processes. The data processing capabilities of Artificial Intelligence which employs algorithms allows for accurate, timely and sophisticated analysis of information superior to human capabilities (BCG 2017: 10). These considerations also lead the turn towards algorithmic systems in public administration, which is described as aimed to satisfy the demand for “greater, objectivity, evidence-based decision-making, and better understanding of individual and collective behaviour and needs” (Lepri et al. 2017: 613).

Apart from the positive connotations presented above, many scholars in the field offer a gloomier view of potential consequences through algorithmic governance. The key points highlighted include privacy concerns, the concept of data culture being antithetic towards delivering public value, selective usage for prior constructed goals, opaqueness of the algorithms, transparency problems and the accountability of far-reaching decisions informed by algorithmic systems (Pencheva et al. 2020; Gamage 2016; Van der Voort et al. 2019; Janssen & Kuk 2016). These issues are in a steady field of tension with the potential benefits that could be derived from harnessing algorithms for informed public sector decision-making.

2.2.2 Big Data

The "fuel" which powers algorithmic governance and essentially all its applications is *Big Data.* Big Data is usually characterized by three typical V-features: volume, variety and velocity. The meaning behind these features are the very large *volumes* of data generated, the very large *heterogeneity* of the data generated and the very *rapid* generation of data (Daniell et al. 2016: 3; Höchtl et al. 2016).

The private sector especially has noted substantial gains and performance increases through the introduction of Big Data systems and the analytical potential which follows (Gamage 2016: 385). But not only the private sector has come to realize the performance enhancement possibilities of Big Data, many further ventures are continuously embracing such technologies and aim to take part in the "modern analytical ecosystem" (Waggoner et al. 2019: 115). Waggoner et al. attribute this rise to a "surge of interest in big data [that] appears to be rooted in [the] desire to leverage millions of terabytes of data generated to understand more substantive phenomena" (2019: 115). Engaging for the public sector is especially the amount of data collected and the drive for the continuous uncovering of phenomena which can help inform policy formulation, implementation and monitoring. Furthermore, what constitutes part of this drive towards leveraging and implementing projects and knowledge through Big Data is the incorporation of unstructured information. This is characterized by information that does not fit the structured spreadsheet norm, but objects "such as email, video, blogs, call center conversations, and social media […]" (TechAmerica Foundation 2012: 10). On top of these mainly communicatory data, governments consistently collect citizen information such as tax records, driving license, medical records, judicial documents, criminal records and many more, which the citizen provides in a steady interaction with the administrative arms of the government. Important to emphasize for the Big Data concept and the usage of predictive analytics in the public sector is the linkage of data. Linking data enables the creation of more complete individual data profiles and may provide deep insights. Through this, expanded analytical applications become feasible as well as the comprehension of complex coherences, which the citizen possibly never intended to provide but which become visible through analytical methods.

Past contributions towards the field of Big Data in the public sector describe the harnessing of key benefits such as effectiveness, efficiency, legitimacy which are to be introduced in the policy making and analysis realm as well as in operational policy aspects (Pencheva et al 2020; Gamage 2016; Höchtl et al. 2016; Katzenbach & Ulbricht 2019; Van Schendel 2019). Höchtl et al. highlights technology as a path towards productivity increases through enhanced information-processing capabilities as well as a reduced time frame and an increased evidence base for informed policy decisions (2016: 148). The turn towards Big Data in policy making is seen by Tsoukias et al. as a way of governing based on facts instead of ideology (2013: 122). Janssen & Kuk further develop this idea of factual dominance in the governance process enabled by Big Data by proposing the possibilities for technocratic governance. Underlying this governance approach is the belief of achieving neutrality by deconstructing the complexity of societal problems into "neatly defined and well-scoped" obstacles in which political realities are not of relevance (2016: 371). On the other hand, the sheer scope of information which is collected and stored allows patterns to be found in Big Data that seem meaningful even though there is no correlation (Mayer-Schönberger 2015: 790). This can lead to highly politicised bits of information with ambiguous interpretations that have contested implications for their further usage. Rob Kitchin has argued that in the policy making process, evidence selection can be skewed and become a specifically agenda-focused activity instead of the technocratic governance exercise envisioned by other researchers (Kitchin 2014 as cited in Poel et al. 2018: 354).

Big Data and its linkage have garnered much attention in the field of governance and policy making as well as for other administrative operations in the public sector. Amidst the data troves becoming ever bigger and more complex, the linkage of information has the potential to enable efficiency gains and support agendas based on the perception of data as a push towards a more neutral and technocratic mode of governance. On the other hand, the possibilities of utilizing Big Data for guiding specific agendas and thus inversing the argument towards "policy-based evidence" also looms large (Strassheim & Kettunen 2014 as citied in Poel et al. 2018: 353).

2.2.3 Predictive Analytics

In this final subchapter, the field of *Predictive Analytics* will be introduced. Making predictions on the outcome of elections, the future unemployment rate and many more phenomena is part of the day-to-day information that citizens, politicians, civil servants, business analysts and many more provide, exchange and absorb.

Predicting something is making "a statement about what you think will happen in the future" (Cambridge Dictionary). Originating from the private sector, the field of business analytics has constantly evolved towards different arms and analytical types. Among these types are description, prediction and decision-making types which the public sector has been keen on incorporating into their operations. Their aim is to leverage the data and apply advanced analytical methods to discover and present meaningful patterns which allow for the quantification, description, prediction and improvement of organizational performance (Daniell et al. 2016: 5). With the computational capabilities intrinsic to algorithms and the large quantities of data, decision-making capabilities have the potential to be greatly improved ex ante. By gauging the computational abilities of predictive systems, data can be searched and analysed many times more efficiently and thoroughly than by human actors. Often times, the output that is generated comes in the form of a risk profile or an indication of the probability that a certain behaviour will set in (Van Schendel 2019: 228-229). Marielle Hildebrandt states:

"The process of 'discovering' correlations between data in databases that can be used to identify and represent a human or nonhuman subject (individual or group) and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category" (2008: 19 as cited in Van Schendel 2019: 227 f.).

This definition of profiling via Big Data and predictive analytics hints towards the concerns associated with this technology. O'Neill (2016) highlights the danger of exacerbating prevailing inequalities by collecting data and making inferences from a group-orientated profile towards individuals (as cited in Lepri et al. 2017: 614). This could lead to self-reinforcing negative outcomes by perpetuating consequences on the individual "based not on their own action but on the actions of others with whom they share some characteristics" (Lepri et al. 2017: 615). Steve McKinlay analyses the debate over the usage of algorithmic systems as one doused and justified in utilitarian terms in which some bias and the presence of false negatives can be regarded as a tolerable trade-off (2020: 155).

As the literature shows, predictive analytics and algorithmic applications in the public sector are strung in a field of tension between computational abilities, efficiency gains as well as negative impacts such as unproportionate profiling and self-reinforcing negative feedback loops. The ambition in some areas of algorithmic administration aim further than the mere replication of present capacities towards the aspired goal of doing "better than humans" (Milner & Berg 2017:15 as cited in Bass 2019: 6).

The three concepts and practices of *Governance by Algorithms, Big Data* and *Predictive Analytics* presented culminate in their relevance and importance for the research question and the theoretical framework which will follow in the next chapter.

These concepts are linked together and bring their own conflicts which result in further complexities and trade-offs when analysing their potential for the public sector. Citizens must see the value that predictive analytics add so that they can conform with the technology. Additionally, they must be able to see benevolent or useful motives in the adoption of algorithms for governing. This stems from a method proposed by Waller & Waller (2020) which examines if predictive analytics are fit-for-use in the public sector through the following attributes: beneficence, non-maleficence, autonomy, justice and explainability.

But trusting the government is not a one-sided process or action stemming from citizens. Governments may understand trust as an outcome which they must drive through actions which positively benefit and reinforce the trust that citizens place in them and their actions. Taking this into account, the second part of this chapter will introduce the concepts of trust and transparency and create the linkage to the concepts presented above.

2.3 Trust and Transparency

2.3.1 Power Asymmetry

The following subchapter will present the *power asymmetry* which could potentially be exacerbated through algorithmic systems entering the public sectors day-to-day operations. There are multiple asymmetries of power which could ensue through Big Data applications and algorithmic systems. This subchapter will focus only on the divide between citizens and the government.

While the key message conveyed on Big Data is its position as an object of knowledge and insight, its role as an object of power must also be taken into consideration (Ruppert et al. 2017 as cited in Vydra & Klievink 2019: 3). This is connected to the neutrality implications that are sometimes presented in relation to Big Data and the belief that "with enough data, the numbers speak for themselves" (Anderson 2008: 7 as cited in Vydra & Klievink 2019: 5). Consequently, some authors present the "aura of objectivity and truth" surrounding algorithms as a de-politicising effect (Boyd & Crawford 2012 as cited in Katzenbach & Ulbricht 2019: 6). A study by Waggoner et al. (2019) presented empirical evidence of this "aura" in action coined "Big Data effect". The authors showed that people envision a connection between quality and Big Data even if they are not able to fully comprehend the underlying processes and functions (2019: 117). Similar to this result, Waller & Waller (2020) provided evidence of the presence of "automation bias" that leads humans "to give greater credence to the outputs of technical systems than their own judgment" (2020: 3). Consequently, the effect presented by Waller & Waller can also run into the opposite direction as "algorithmic aversion".

To fully understand the outputs of algorithmic systems a great deal of data literacy is required for those impacted by the outputs: the citizens. (Poel et al. 2018: 362). Algorithms are complex in nature and through this some are inherently opaque (Poel et al. 2018: 374). This manifests itself in the inability to monitor which data has been collected and what its usage will be, thus muddying the balance between utility and disutility of the application for citizens (Janssen & Kuk 2016: 373). Waller & Waller (2020) emphasize that for algorithms to be used ethically, transparency and explainability are essential (2020: 9). If the collected and used data is not monitored and the necessary literacy to comprehend the outputs is missing, it will become harder to distinguish if a policy is driven by evidence or if the evidence is driven by the policy (Strassheim & Kettunen 2014 as cited in Poel et al. 2018: 253).

In public sector usage this point must be especially emphasized, as citizens cannot opt-out of government and therefore have little direct control to how their data is processed in such systems. Apart from these fields of tensions, technological advancements in the past have also been critically perceived. The late sociologist Heinrich Popitz stated that these advancements can increase the capacity of governments to gather knowledge and thus exert power (1992: 181 as cited in Peters & Schuilenberg 2018: 269). John Wanna also regards advancements in algorithmic governance as a further shift of power asymmetries between governments and civilians (2018: 6). Formulating the potential for malevolence in a more concrete fashion, Katzenbach & Ulbricht (2019) warn of algorithm-based valuation practices which "[…] create stratification mechanism[s] that can superimpose social class and reconfigure power relations often to [the] detriment of the poor and underscored" (2019: 5).

As a conclusion, it becomes clear that opposed to the efficiency gains and perceived neutrality of algorithmic systems, the potential for power asymmetries between governments and its citizens is also an incremental part of the debate, leading to the first hypothesis:

H1: *Citizens that perceive predictive analytics as instruments which widen the power asymmetry between themselves and the government will exhibit a lower score of trust in government.*

Citizens focusing on the potential of predictive analytics to make inferences about individuals via data which was not provided and thus fear a more intrusive and surveillance-orientated state will trust the government less. Due to the nature of government and the inability to opt-out, citizens reaction to a perception of widening power gap can be a withdrawal of trust not only in the technology but in the government as a whole. Therefore, citizens must be able to fully understand the utilities and disutilities of why the government is pursuing the implementation of predictive analytics. I thus assume, that a perception of widening power asymmetry correlates with a lower trust score.

I understand the power asymmetry as a process or state of the relationship that is not necessarily intentional on behalf of the government but may result through the combination of Big Data and predictive analytics. Thus, while governments may be unaware of a widening power asymmetry, citizens can perceive it. This is further clarified in the research design and included in the survey to capture and analyse the perception of the respondents. Furthermore, I propose that the tool to foster understanding and acceptance for the citizens may be transparency, which is decided and provided by the government.

The next chapters will focus on algorithmic aversion, the role of transparency in algorithmic systems and how transparency can act as an intermediary of trust in government.

2.3.2. Algorithmic Aversion

Apart from a fissure in the trust relationship between citizens and their governments, citizens may also project mistrust towards the technology behind predictive analytics: algorithms. This reaction is called algorithmic aversion and was introduced by Dietvorst et al. (2015) who demonstrated that even though algorithmic forecasts in their specific experiment were more accurate than those of human forecasters, the respondents placed higher confidence in the humans (2015: 2). This effect was further increased when the respondents were able to witness the algorithms make an error. The respondents were less tolerable towards algorithmic mistakes than towards human forecasters confirming their hypothesis that "people are quicker to abandon algorithms that make mistakes than to abandon humans that make mistakes, even though, as is often the case, the humans' mistakes were larger (2015: 6). Furthermore, their study suggests that people's willingness to rely on algorithmic forecasters is given when they cannot see them make the error, when the "errors are unseen, the algorithm is unseen […], or when predictions are nearly perfect" (2015: 11). Dietvorst et al. point out that while most people generally are more comfortable around simple and transparent algorithms, there is still research to be done as to which attributes further contribute to algorithmic aversion (2015: 11).

The concept of algorithmic aversion is intriguing for discussions surrounding algorithms in the public sector. Most democracies tend to act in such a way that their citizens overall understand and accept administrative actions. Following the hypothesis of Dietvorst et al., predictive analytics for example would receive a lower margin for error by citizens who would expect nearly perfect functioning. Furthermore, the areas of application in the public sector often require vast amounts of data to be collected from multiple sources and in different formats which raises the complexity of the algorithm thus lowering explainability and transparency. The high-stake decisions in which predictive analytics may be involved in the public sector such as decisions on parole, the granting of welfare benefits and asylum admissions may further reduce citizens acceptance of errors and lead to an overall aversion of algorithms hindering the technological development within the sector. Furthermore, I assume that those impacted by and sceptical of algorithmic applications will not only exhibit a negative reaction towards the technology, but also towards the employer of the technology. Thus, taking Dietvorst's concept into account, the second hypothesis relates to how acceptable citizens deem predictive analytics for the decision presented in the hypothetical scenario:

H2: Respondents from the least transparent scenario will find the usage of predictive analytics for far-reaching decisions less acceptable than those from the most transparent scenario.

This hypothesis is directly related to the second research question which analyses citizens attitudes towards a government that uses predictive analytics. I assume, that the respondents that were exposed to the scenario with the least transparent vignettes will also be more inclined to deem far-reaching decisions that are based on an algorithmic output to be unacceptable than those who were exposed to the most transparent vignettes.

2.3.3 Transparency and Algorithmic Systems

To clearly present how transparency and opacity are omni-present around algorithms the following subchapter is further divided into two parts. The first part will present the notions of transparency and opacity which are intrinsic to algorithmic applications and are often circumscribed as "black-box systems" in the literature. The second part will build up on the intrinsic characteristics presented and analyse transparency and opacity surrounding the implementation and operation of algorithms.

Forming a collective understanding of transparency for this thesis, the definition proposed by Stephan Grimmelikhuijsen et al. will be used. In the authors work on governmental transparency, they relate the definition “to an entity’s revelation of information about its own decision processes, procedures, functioning and performance” (Grimmelikhuijsen et al. 2013: 3-4). Furthermore, they add additional incorporated components “including inward observability, active disclosure and external assessability” (Grimmelikhuijsen et al. 2013: 4). Grimmelikhuijsen et al.’s definition and understanding of transparency was chosen due the linkage towards the directions and varieties of transparency presented by Hood and Heald (2006) which are an integral part of the research design for this thesis. The revelation of information on an entity’s own decision coincides with Hood and Heald’s upward direction of transparency, in which agents monitor the principle. This is only possible if information is revealed by the entity itself. Secondly, the inward observability coincides with the inward direction of being able to actively monitor an organization or entity’s innerworkings and procedures. Active disclosure covers the three effective varieties of transparency as proposed by Hood and Heald: process, real-time and effective transparency. These three varieties enable an understanding of active disclosure and external assessability. Opposing transparency is opacity. Burrell (2016) defines three different types of opacity related to algorithms: intentional, illiterate opacity and intrinsic opacity (Burrell 2016 as cited in Lepri et al. 2017: 619). The following two subdivisions will present first intrinsic opacity and after extrinsic opacity and its relevance for transparency and trust.

Intrinsic Opacity

Among the challenges faced by algorithmic systems such as predictive analytics and their usage in the public sector is their inherent complexity. Some authors describe this characteristic as that of a “black-box”, in which the outcome cannot be traced back to the processing that was done within in the system. Approaching black-box systems from an engineering perspective where only the performance of the task matters regardless of how it was completed within the black-box differs greatly from the multi-stakeholder perspective taken on in the public sector (Citron & Pasquale 2014: 6). Furthermore, the intentional and intrinsic opacity types provided in the framework of Burrell (2016) are relevant components of assessing algorithms for public sector usage and are complemented by extrinsic opacity.

Burrell describes intentional opacity as an objective for protecting the intellectual property rights of the inventors of the algorithm (Burrell 2016 as cited in Lepri et al. 2017: 619).

Depending on where public administrations and governments source their algorithmic technology, this could be subject to intellectual property protection on which the supplier may insist. Consequently, this could lead to citizens as well as civil servants unable to assess the inner-workings of the system on which they base their decision, leaving them only with the output. Lepri et al. acknowledge that legislation which forces decision-makers to use open-source systems could mitigate this form of opacity through mechanisms such as the EU General Data Protection Regulation (GDPRs) "right to an explanation" but on the other hand this could hamper the rollout of such technologies due to a lack of compliant suppliers (2017: 620). The second type, intrinsic opacity, deals with the notion of black-box systems as described above. Problematic is mainly the interpretability of the underlying methods of the application. This can be counteracted by using alternative models that are easier to interpret but therefore often yielding a reduction in accuracy giving way to more trade-offs that need to be considered (Lepri et al. 2017: 620).

Extrinsic Opacity

Opaqueness may also be brought about during the implementation and operation of algorithms coinciding with the third opacity type mentioned by Burrell (2016): illiterate opacity. Due to the high degree of technical understanding needed to grasp the procedures, functioning and underlying technology of algorithms, the lack thereof can also be factor of opaqueness (Lepri et al. 2017: 619). By lacking algorithmic and technical literacy, citizens as well as decision-makers are unable to shape informed perspectives on the utility or disutility of the application complicating its usage for public sectors in Democracies. Additionally, the "right-to-explain" under the GDPR is somewhat diluted if those who are responsible for the explaining have difficulties with the matter as well as those receiving the explanations.

Both the intrinsic as well as the extrinsic opacity present with algorithmic systems provides for an intriguing basis to analyse how transparency could counteract this and maintain trust in governments using these applications. Former U.S. Supreme Court Judge Louise Brandeis stated that "sunlight is the best disinfectant". But how can a system with inherent opacity nonetheless be used to predict events and make inferences on individuals without eroding trust in the government? And how is transparency more generally important for trust in government?

2.3.4 Directions, Varieties and Outcomes of Transparency

While it is established that citizens are not able to opt out of government, they have other possibilities to signal and express their discontent. Lind (2018) describes citizens as consumers of government which instead of opting out, have the possibility to withdraw their trust and through this end their engagement with the government (2018: 92). Consequently, it is incremental for governments and their public administrations to engage with citizens in such a way that their trust is retained. This following chapter will present the concepts of trust and transparency and their transferability to predictive analytics and the main research question.

In the public sector, transparency can be described as a value and as a vehicle for beneficial outcomes and constructive relationships. Furthermore, it is both a principle and a practice:

"As a principle, it highlights what a government agency should aspire to do in order to become transparent, and as a practice, it is the specific actions that enable a government agency to become more transparent to its employees and external stakeholders through, for instance, improved communication." (Burke & Teller 2011 as cited in Graham Stone 2016: 571).

By improving communication and enabling the sharing of information, transparency becomes an incremental aspect of ensuring the reliability and integrity of public institutions, leading to stronger public trust and higher support (Jashari & Pepaj 2018: 61). On the other hand, sceptics such as O'Neill (2002) believe that transparency in a too high "dosage" may achieve the opposite effect and erode trust and undermine governance (Meijer et al. 2018: 501). Furthermore, Meijer et al. (2018) argue that the specific context in which transparency is channelled varies greatly and thus also the reception by citizens may differ significantly (2018: 502). They point to divides that exist between citizens such as levels of education and information savviness.

But to fully understand the innerworkings of transparency it is important to emphasize the three elements which characterize transparency. Foremost, there must be an observer, some piece of information or a process which is available to be observed and finally the existence of means or a method for observation (Oliver 2004: 2 as cited in Meijer 2013: 430). In the case of predictive analytics, the intrinsic technological complexity may potentially obstruct the creation and sustainability of transparency with unknown effects and reactions from citizens.

Apart from the three elements needed for transparency, there are also multiple varieties of transparency as well as directions through which transparency works. Hood and Heald (2006) describe these four different directions: *transparency upwards*, which exhibits a hierarchical structure and where the principle monitors the agent*; transparency downwards*, in which the agent may access information and monitor the principle; *transparency outwards*, in which an agent is able to observe the outside of an organization and finally and *transparency inwards*, in which observers can evoke the innerworkings of the an organization (Hood and Heald 2006 as cited in Rodrigues 2018: 240).

In the case of predictive analytics, I assume that citizens will have greater trouble accessing transparency upwards as well as inwards as the inherent opacity of the algorithm constitutes a natural barrier for most citizens to understand the workings and thus make use of provided information. Additionally, this technology which is characterized by its high computational and analytical ability poses a new method of governments exercising transparency downwards and thus monitoring the citizen coinciding with an increased power asymmetry. I assume outward transparency to remain a more neutral direction which is not greatly impacted by the usage of predictive analytics.

Apart from the directions of transparency, Hood and Heald (2006) emphasize multiple relevant varieties of transparency. They propose three different varieties: *event* vs. *process* transparency, transparency in *retrospect* vs. transparency in *real-time* and *nominal* vs. *effective* transparency (Hood & Heald 2006 as cited in Rodrigues 2018: 240).

The authors stipulate that gaining transparency on events is a simpler undertaking than enabling transparent processes due to the differences in the measurability of an event as a policy point than the various measures which accompany policy implementation (Rodrigues 2018: 241). In this case, I propose taking both event as well as process transparency into account instead of seeing them as opposed varieties. Hood and Heald state that event transparency is simpler to achieve than process transparency. This follows from processes being related to the various measures that are taken to implement a policy (Rodrigues 2018: 241). Governments using predictive analytics will encounter singular cases where transparency is relevant as well as transparency accompanying the process of procurement, implementation and operation. Thus, in the context presented in this thesis I take both varieties of process and event transparency into account. Secondly, transparency in retrospect may be a tool to avoid breaching confidential information and thus protect the policy, the government as well as other relevant actors such as the citizens. On the other hand, a delay in the flow of information may also be problematic. Real-time transparency enables a continuous monitoring of the information either of the whole process or at specific points (Rodrigues 2018: 241). Finally, Hood and Heald distinguish nominal and effective transparency. This differentiation allows for assessing if the disclosed information is done in a simple manner and for the sake of disclosure or if the information is processed so that actors such as citizens may scrutinize and understand the data in an informed manner (Rodrigues 2018: 241). Nominal transparency coincides with Grimmelikhuijsen's concept of pseudo-transparency, in which organizations appear to be transparent by providing great swathes of information (2012: 53). Furthermore, Grimmelikhuijsen and Welch (2012) emphasize the differences between policy content and policy outcome transparency. While policy content transparent is related to information disclosed about the policy itself and accompanying measures, their problem-solving ambitions, method of implementation as well as the implications for citizens and other stakeholders, policy outcome transparency "captures the provision and timeliness of information about the effects of policies (2012: 4).

Applying these theoretical blocks to predictive analytics, the difficulties facing governments willing to adhere to transparent principles as well as citizens which expect transparency in a meaningful and timely manner become clearer and allows for the formulation of the third hypothesis:

H3: *Transparency, in the usage of predictive analytics, positively affects reported trust in government if it is directed upwards as well as inwards and is additionally characterized by the varieties: event & process, real-time and effective transparency.*

By proposing these conditions in the hypothesis, I aim to construct a "most transparent" scenario with the directions and varieties presented by Hood and Heald. I propose the following relationship between the directions and varieties in relation to predictive analytics. The directions of upward and inward would not only enable citizens to access the rationale behind the government's usage of predictive analytics but also monitor the innerworkings of the specific organization using the technology. Furthermore, by presenting itself transparent in both singular events as well as the process as a whole, the government may foster further trust by its citizens. By adhering to real-time transparency, citizens remain in the loop as to where, for what and on whom the government is using predictive analytics. By fulfilling the notion of effective transparency, the information provided by the government is processed in a meaningful manner that citizens are able to understand it and react in an informed fashion.

Transparency as a practice, principle and value is usually not adhered to just for its own sake but also to achieve or drive specific outcomes. John Wanna considers multiple outcomes to transparency such as transparency as a virtue to which governments and public administrations can aspire, an effective and efficient enabler allowing citizens to benefit from transparent rules and information, an improver of accountability through public scrutiny and insight and as a promoter of confidence and assurance which aims to maintain trust in public institutions (2018: 12). In a similar fashion, Cucciniello et al. (2017) also emphasize transparency as a means of achieving objectives such as improving performance or fostering trust (2017: 32). Deeming as trust is an essential building block for governments to retain support and achieve meaningful relationships with their citizens a definition for this context is needed: Alessandro et al. (2020) define trust in the public policy sphere as a "psychological state involving positive confident expectations about the competence, benevolence, honesty and predictability of another person or organization, and the willingness to act based on these expectations" (2020: 2).

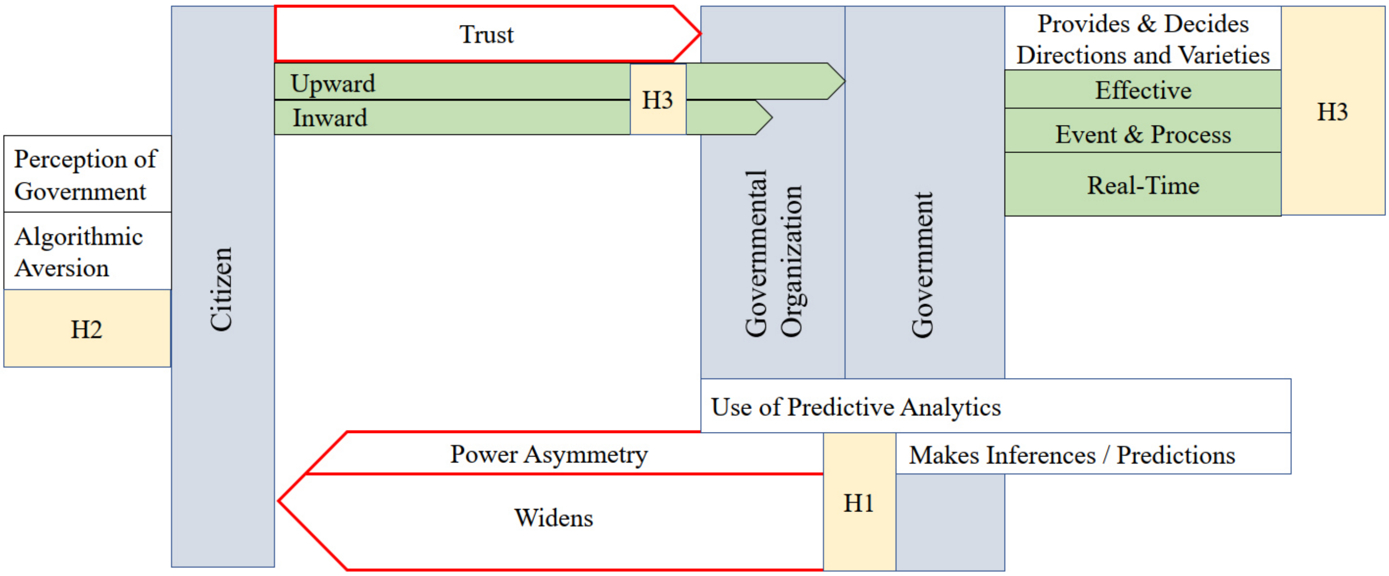
Predictive analytics and other algorithmic applications are posed to have a large impact in the public sector. Looking back at the previous chapter's multiple trade-offs und tension fields arise such as between privacy and efficiency, differing levels of data and technological literacy and the intrinsic characteristics of predictive analytics that potentially stand opposed to the value of transparency highlight some of the difficulties surrounding this topic.

As presented, transparency is one of the intermediaries for governments to not only strengthen their legitimacy and prove their competence but also to achieve outcomes such as higher trust. On the other hand, governments find themselves between "a rock and a hard place" with the implementation of predictive analytics. Not only may this technology intensify a power asymmetry between governments and their citizens, but the vast analytical power allows governments to make inferences and predictions of their citizens with data the citizens never intended to provide.

Due to the very nature of government, opting-out is impossible but citizens may revoke their trust and through this withdraw their support. All in all, for governments to be able to both achieve the desired efficiency gains through predictive analytics while still maintaining the relationship of trust with their citizens they will most likely have to channel this through an intermediary such as transparency. As described, transparency in the realm of predictive analytics is especially intriguing due to its characteristics as well as the directional changes of transparency it could perpetuate. Furthermore, the different varieties of transparency will most likely play a large role in how much trust governments will generate, retain or lose.

2.3.5 Conceptual Framework

To conclude this chapter the conceptual framework will be presented which connects the theoretical assumptions, concepts and hypotheses to present the basis for the research design and methodology. To further clarify the relationships, a visualization is introduced.



Conceptual Model based on Hood & Heald (2006); Own Visualization

The *government*, which is the *provider and decider of directions and varieties* of transparency is made up of *governmental organizations.* Building upon the most transparent scenario as described in chapter 2.3.4, the citizens may monitor the government (*upward*) as well as the innerworkings of governmental organizations (*inward*). Furthermore, in this visualization the government provides the varieties to achieve the most transparent case: *effective, event & process and real-time* transparency.

The government and more specifically the governmental organization employs *predictive analytics* which *makes inferences/predictions* on events or citizens behaviour. Using predictive analytics may also *widen* the existing *power asymmetry* between citizens and the government. This is also where the first hypothesis (*H1*) is situated, that citizens that perceive predictive analytics to widen the power asymmetry also exhibit less trust in the government. The second hypothesis (*H2*) is situated by the citizens and encompasses their perceptions and more specific reactions such as algorithmic aversion. Finally, the third hypothesis encompasses the upward and inward directions as well as the three positive varieties and states that if these directions and varieties are present, it positively affects citizens trust in government in the usage of predictive analytics.

3. Research Design

In this chapter I will present the *Research Design* which was used to gather, format and analyse the data. The first subchapter presents the *Methodological Approach* and argues why the chosen approach is best suited to answer the research questions of this thesis. Subsequently, the subchapter *Data Collection* will be presented which includes specifics related to the sampling method and the procedure of the data collection. The third subchapter *Data Analysis* describes the process of data preparation and statistical analysis including a brief section on the reasons for the specific analytical choices.

3.1 Methodological Approach

To commence this chapter, the two main research questions will be reiterated before the methodological approach to answer them is explained more specifically:

RQ1: *How does the application of predictive analytics affect the relationship of trust between citizens and the government?*

With the first research question, I would like to analyse if the usage of predictive analytics leads to a change in the relationship of trust between citizens and their government. Deeming as governments have many functions and each individual may include a multitude of different factors into his or her perception the main focus point for this methodological approach is to isolate the government as the user of predictive analytics and the citizen who is analysed. The first hypothesis as well as an open-ended question, asking respondents why their perception of the government changed, are aligned with this question.

RQ2: *Does transparency influence citizens attitudes towards a government that uses predictive analytics?*

With the second research question, I set out to examine if transparency may influence this relationship, building on the directions and varieties of transparency as proposed by Hood and Heald (2006). The second as well as the third hypothesis and answers from the open-ended questions are related to this sub-question.

While there are multiple stand-out surveys which measure perceptions of trust such as the World Values Survey and the Edelman Trust Barometer, the data collection for this thesis relies on primary sources to be able to focus on the specific content of the research interest. The data is collected via a survey design that includes distinct vignettes within two hypothetical scenarios. I chose this approach to exert full control over the differences that the respondents encounter between the different scenarios. By only making the distinct vignettes for the directions and varieties of transparency and keeping the follow-up questions identical, I aim to find differences between the answers of respondents that were allocated to different scenarios while maintaining comparability. The scenarios have varying vignettes which change the level of transparency present in the specific scenario building up on Hood and Heald (2006) as shown in Chapter 2.3.4. More specifically, the directions and varieties were split between positive and negative transparency allowing for a *most transparent* and *least transparent* scenario.

The *most transparent* scenario includes five positive transparency vignettes: upward, inward, event & process, real-time and effective transparency. The *least transparent* scenario on the other hand included three negative and one neutral transparency vignette. Included in the negative vignettes is downward, retrospect and nominal transparency. Outward transparency was included as a neutral direction.

Directions and Varieties:		
Least Transparent		**Most Transparent**
Downward (-)	vs	Upward (+)
Outward (=)	vs	Inward (+)
Retrospect (-)	vs	Real-Time (+)
Nominal (-)	vs	Effective (+)
		Event & Process (+)

Table 1: Directions and Varieties of Transparency

The basis of the scenario is a hypothetical situation in which a citizen in a democratic state has his welfare benefits halted due to his case being marked as potentially fraudulent. While this starting point is the same for both scenarios, depending on whether it is the most or the least transparent scenario the actions and attributes of the government vary and through this the levels of transparency. By describing this scenario, each respondent is exposed to a similar stimulus but with slightly varying characteristics. This allows the respondents to judge these specific stimuli allowing the survey design to profit from a relatively high internal validity due to the experimental-type setting (Auspurg & Hinz 2006: 7).

While there are concrete examples of the usage of predictive analytics in the real-world, the technology is not routinely used or discussed. This allows the survey to create a clear situation which respondents use as an anchor point for the follow-up questions. In this case, a vignette survey study enables the researcher to "expand reality" and create "[…] a hypothetical description of situations or subjects for evaluation […] of stimuli that do not (yet) exist in reality" (Auspurg & Hinz 2006: 8).

As the basis of the scenario in this survey has similarities to the SyRI programme in the Netherlands and a programme for allocating the unemployed to active labour market programmes in Poland, I specifically avoided mentioning a country or tying the scenario to one of these examples for multiple reasons. Foremost, by choosing a specific country and then examining how predictive analytics could impact trust between citizens and their governments, the problem of confounding variables and high levels of variation in respondents' answers looms large. Respondents may generally mistrust or trust their government due to the party that is in power or other effects which could directly influence how much trust they place in the usage of predictive analytics. Secondly, by emphasising a past project such as SyRI in the survey the same problem may arise. Respondents that are informed on this specific case and its outcome, may bias the research by tying their perception of predictive analytics to a specific past case.

Directions and Varieties of Transparency: Least Transparent			Most Transparent	
Downward	[…] is keen on understanding its citizens and collects and analyses broad swathes of data including with algorithms and advanced analytical applications	vs	**Upward**	[…] publishes widespread, accessible information on governmental activities and actively encourages citizens to scrutinize this information
Outward	The citizen has had contact with the organization during the application process for the benefits, but due to the little information they publish on their procedures and innerworkings, the citizen feels unfamiliar with the organization	vs	**Inward**	The citizen is familiar with the governmental organization in charge of handling the benefits due to the clear instructions provided by the caseworker at the organization as well as through their quarterly updates
Nominal	The citizen receives a document clarifying that all data provided to the Government may be used as an input to the algorithm. Furthermore, the organization states that they may not publish the exact innerworkings of the algorithm to avoid third parties from "gaming" the system and avoiding detection	vs	**Effective**	The answer the citizen receives from the inquiry is an explanation as to which data is collected and used as in input of the algorithm. This led to the case showing similarities to other fraud cases
Retrospect	The answer the citizen receives from his inquiry is that the case has been analysed using a predictive algorithm which assessed the data provided and concluded that there are similarities to past cases of fraud	vs	**Realtime & Event**	Furthermore, in the letter the citizen received, the organization gives notice that the halting of the decision to discontinue the benefits payment is based upon the output of an algorithm used to analyse data and make predictions
			Process	[…] the Government has decided to utilise predictive analytics to make full use of the data through more efficient analysis and the possibility of predictions. Both where and why this technology will be used is communicated by the Government

Table 2: Survey Vignettes

3.2 Data Collection

This subchapter will further clarify the means of data collection, the procedure and give insights into the type of questions included in the survey as well as the reasoning behind them.

As stated above, the data was collected via a vignette survey approach. The survey was published on 03.05.2021 and was active until 14.05.2021. QualtricsXM was used and the survey was shared via a direct, anonymous link with personal contacts as well as a Bachelor's level course from Leiden University. The link was re-sharable, thus enabling participants to further distribute the survey. Upon deactivation of the survey, there were n= 69 respondents that completed the survey, distributed alternately to either the most transparent or least transparent scenario. In total 30 respondents started the survey but dropped out during the process. These partial responses were stored for 7 days to allow respondents to complete the survey and then automatically deleted. All surveys which were started and not completed were deleted upon deactivation of the survey. Further 7 respondents were excluded as they did not meet the requirement of passing an attention check question, leaving n=62 respondents for the final analysis. The distribution was randomised in Qualtrics and was determined once the individual respondent opened the link. All multiple-choice questions were constructed on a five-point Likert-scale and numerically pre-coded with the Qualtrics programme. Taking the debate into account if the Likert-scale should include a middle point such as neutral, no change, or neither agree nor disagree, I decided to include the middle option (Chyung et al. 2017: 4-5). While this may give rise to central tendency bias and respondents avoiding the "extreme" options, leaving a middle point out could lead to acquiescence bias and respondents agreeing or disagreeing with positions that they feel neutral about or have no opinion on (Taherdoost 2019: 4). Seeing as respondents may feel indifferent about the usage of predictive analytics and their perceptions of the government in this case, I included the middle option and believe it is an important addition to achieve the most balanced and thought-through responses while allowing me to measure neutral as well as "extreme" positions.

3.2.1 Basic Questions and Vignettes

The survey is partitioned into four different blocks. The first block shortly presents the research interest and the informed consent paragraph. In this part, the research interest is shown only on a general level without clarifying specifically that transparency and trust are the main objects of research. I decided to avoid mentioning these concepts as respondents may exhibit reactions to positioning themselves as trusting or mistrusting the government. This reasoning was further motivation to choose the vignette experiment, as "[…] it allows the subject to avoid directly admitting holding a controversial position or behaviour while still allowing the researcher to make inferences" (Guy Peters & Guedes-Neto 2020: 224).

The second block of the survey consists of four basic questions, asking for age, gender, tech-savviness and privacy concerns related to data sharing. By asking for perceived tech-savviness, I aim to analyse if there are discrepancies between respondents from the different scenarios which require further control or attention. The question originates from a study on the usage of mobile phones, privacy and security practices by Krskovsky and Syta (2010). To further clarify how tech-savviness is perceived in this question, a table was added in the survey explaining and anchoring three out of five points on the Likert-scaled answer.

Following the basic questions, respondents are presented with the third block, that consists of a random allocation to one of the two hypothetical scenarios.

3.2.2 Transparency, Trust and the Power Asymmetry

The fourth block contains the follow-up questions to the hypothetical scenario and is the main instrument to collect the necessary data. There are twelve questions presented in a multiple-choice manner on a five-point Likert-scale as well as two questions with an open text box response to give the respondents the possibility to further substantiate an answer. Following survey research practice, the follow-up questions were developed in such a manner that the same construct is measured through multiple lenses (Andres 2012: 3).

The first two questions measure transparency and its effectiveness in the specific scenario presented to the respondent. By using the proxies for transparency in both questions, I expect the respondent to think about the interaction in the specific scenario he or she received and avoid bringing individual sentiments towards governmental transparency into the response. The first proxy question presents the respondent with a statement if the citizen in the scenario has been sufficiently informed by the government over the actions taken. The second proxy asks respondents if the government has acted in such a way that the citizen has the possibility to understand the decision. This question aims to analyse if differences arise between respondents of the scenarios. Relating to the third hypothesis, I expect respondents that were allocated to the most transparent scenario to exhibit higher values on the transparency score due to the vignettes they were exposed to that encompass the "positive" directions and varieties of transparency.

The following three questions are constructed as proxies of trust. The first question asks the participants if they believe that the citizen will feel uneasy with sharing data in the future. Feeling uneasy or becoming more cautious has been used in a past survey on trust by Miller and Mitamura (2003). The authors explained that there must be a conceptual distinction between trust and caution as it is possible to trust people or institutions while at the same time believing in the prudence of cautiousness (Miller & Mitamura 2003: 63). In the question presented in this survey, the cautiousness or feeling of unease in future data sharing with the government enables me to measure generally how citizens would interact in the future with governments that use predictive analytics. The second question in the block asks if predicting individual's behaviour may lead to withdrawal of support on the citizens side and was also constructed as a proxy of trust. The third question continues the hypothetical scenario and asks the respondent to imagine that the citizen did not commit fraud but was labelled as such by the algorithm and if this would lead to the citizen having less faith in the competence of the government. Competence as a proxy stems from the definition of trust in the public policy sphere as defined by Alessandro et al. (2020) in Chapter 2.3.4.

Next, a question is posed asking the respondent to assess the acceptability of basing a decision with far-reaching consequences on the output of algorithm. This question is inspired by Dietvorst et al.'s (2015) study covering algorithmic aversion. This concept is placed as an attitude in the conceptual model with which citizens may react towards a government that uses predictive analytics. While the core reasoning behind algorithmic aversion is the mistrust or rejection of an algorithmic decision maker, I assume that citizens who believe the decision in the scenario is unacceptable will also have an attitudinal change towards the government employing predictive analytics and I include this as an integral part of my second hypothesis.

Before the block assessing the power asymmetry begins, respondents face an attention check question which asks about the content of the hypothetical scenario. Those who are not able to correctly answer the question are omitted from the final analysis, as it is unsure if their answers are placed randomly.

The next block of questions is based on the power asymmetry explained in Chapter 2.3.1. Three questions are included, and the content is based upon statements which are derived from the hypothetical scenario. The first question assesses respondent's opinion on the possibility that governments employing predictive analytics may have greater chances to surveil citizens even if this is not intended. Secondly, I ask respondents to assess if information on individual citizens becomes more visible for the government. The final question in the power asymmetry block asks the respondents if the possibility to link and analyse data may lead to privacy infringements of individuals. By combining these questions to a joint power asymmetry variable, I aim to test the first hypothesis in which I expect citizens that perceive this asymmetry to be widening, will also exhibit a lower score of trust. These proxy questions are also related to statements presented in the theoretical chapter such as from Heinrich Popitz who stated that technological advancements my increase knowledge gathering capacities on the government side which could lead to a larger exertion of power (1992: 181 as cited in Peters & Schuilenberg 2018: 269). Furthermore, the second proxy builds on my assumption that the widening power asymmetry is not necessarily intentional from the governments side.

At the end of the survey, I ask respondents to imagine themselves in the position of the citizen and assess if their future perception of the government would change. This allows me to assess if the respondents perceive the actions of the government and the usage of predictive analytics in a more benevolent or malevolent fashion. To give respondents the chance to explain why their perception would change, a text box is added in which comments can be added.

The last question of the survey asks respondents if they believe that predictive analytics may be a useful tool for governments. This question is also followed up with a text box in which respondents may explain their assessment and allows me to analyse how citizens generally view the utilities or disutilities of predictive analytics and which factors influence their view. I believe this is an important question to analyse further concerns or assumptions about the usage of the technology in the public sector.

3.3 Data Analysis

All respondent’s data was collected anonymously via the Qualtrics XM survey programme. To enable the possibility of sorting respondents by the hypothetical scenario they were assigned to, each scenario was further equipped with a distinct single choice question.

Hypothetical Scenario a.), which was the most transparent scenario included the *Proceed to follow-up questions* while Hypothetical Scenario b.), the least transparent scenario included *Continue to follow-up questions* as a response item. This allowed me to create a distinct filter for each scenario, enabling a comparison of answers from the most to the least transparent scenario.

The collected data was then exported to SPSS for further analysis. First, new variables were created from the items that the questions measured. Three variables were created: TransparencyC, which presented both transparency items combined, TrustC, presenting the three trust variables combined and PowerC, which included the three items testing for the perception of a power asymmetry. The variables were computed by adding their median values together. The median was chosen instead of the mean due to the ordinal scale level that is inherent to Likert-scaled data (Chyung et al. 2017: 3). While some authors opt for combining Likert-scaled items and treating them as interval scale data, I believe that in this case continuously treating all items as ordinal data is the suitable approach. While this alters the statistical methods, which can be employed in the following analysis, assuming that respondents equally weight the distance from *agree* to *strongly agree* and thus assuming interval-scaled data “may increase the chance of coming to the wrong conclusion about the significance (or otherwise) […]” of the research (Jamieson 2005: 2).

After computing the new variables, I conducted a reliability test using Cronbach’s α as a measure.

Reliability of Items:	N	Items	Cronbachs Alpha
Transparency Combined	62	2	0,732
Trust Combined	62	3	0,604
Power Asymmetry Combined	62	3	0,518

Table 3: Item Reliability

The item measuring transparency showed a Cronbach's α of 0,732 indicating a reliable scale. The trust and power asymmetry items scored lower but were still in the threshold of a moderately reliable scale (Hinton et al. 2004: 363). To further assess the reliability of the combined variables, I assessed the inter-correlations between the items. None of the items exhibited a negative correlation allowing me to assume that the items have been correctly scored and measure the same characteristic although in some cases on a weak level (Pallant 2016).

Following the reliability test of the items, I conducted normality tests to analyse the distribution and assess which statistical methods are applicable for further analysis. The three variables tested were the combined items for trust, power asymmetry and transparency using both the Kolmogorv-Smirnov as well as the Shapiro-Wilk test.

Tests of Normality:	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Trust	0,294	62	0,000	0,761	62	0,000
Power Asymmetry	0,299	62	0,000	0,804	62	0,000
Transparency	0,175	62	0,000	0,9	62	0,000

Table 4: Tests of Normality

All variables were statistically significant, leading me to reject the Nullhypothesis and view the variables as not normally distributed. Having clarified this and deeming as all variables are ordinally scaled, the statistical tests which are possible for the data present are non-parametric tests such as Spearman's Rho.

4. Results

The following chapter will present the empirical findings generated through the survey. First, I will present descriptive statistics showing the basic demographics of the participant pool such as age and gender. Following this, I will analyse the questions in their grouped form as elaborated in Chapter 3 and compare the results of the most transparent and least transparent scenario. Finally, the results are discussed and the hypotheses either accepted or rejected.

4.1 Descriptive Statistics

The following table will present some basic descriptive statistics and frequencies of the survey respondents. The amount of male and female respondents was relatively balanced, with 29 male respondents and 32 female respondents. One participant identified themselves in the non-binary / third gender category and no respondents opted to withhold their gender via the *prefer not to say* option. Furthermore, the two dominant age cohorts are those ranging from 18 – 24 years and 55 – 64 years with 35,5% and 30,6% respectively.

Gender:	Frequency	Percent	Cumulative Percent
Male	29	46,8	46,8
Female	32	51,6	98,4
Non-binary / third gender	1	1,6	100,0
Total:	62	100,0	
Age:			
18 - 24	22	35,5	35,5
25 - 34	11	17,7	53,2
35 - 44	3	4,8	58,1
45 - 54	4	6,5	64,5
55 - 64	19	30,6	95,2
65 +	3	4,8	100,0
Total	62	100,0	

Table 5: Descriptives: Gender & Age

54,8% of respondents indicated that they have an *intermediate* level of tech-savviness, with no respondents indicating a *low* score. Only four respondents indicated that they situate themselves at the highest tech-savviness level*: advanced*. Reporting on the question asking for privacy concerns related to personal, the data exhibited more variation in the responses. 54,8% or 34 respondents indicated that they are *somewhat concerned* about privacy matters when providing personal data. No respondents answered that they are *completely unconcerned* and only six respondents feel *somewhat unconcerned*. The highest indication *very concerned* was selected by 22,8% or 14 respondents.

Tech-savviness:	Frequency	Percent	Cumulative Percent
Low/Intermediate	4	6,5	6,5
Intermediate	34	54,8	61,3
Advanced/Intermediate	20	32,3	93,5
Advanced	4	6,5	100
Total	62	100	
Privacy Concerns related to Personal Data:			
Very concerned	14	22,6	22,6
Somewhat concerned	34	54,8	77,4
Neither concerned nor unconcerned	8	12,9	90,3
Somewhat unconcerned	6	9,7	100
Total	62	100	

Table 6: Descriptives: Tech-Savviness & Privacy Concerns

The respondents were allocated alternately to either the most transparent or the least transparent vignette upon opening the survey window. Due to respondents dropping out after having started the survey, the final distribution is uneven. 27 respondents were allocated to the most transparent scenario and 35 respondents were allocated to the least transparent scenario. In both scenarios, the age group ranging from 18 – 24 years is the largest with 11 respondents. The least transparent group has nine respondents ranging from 25 – 34 years while the most transparent group only has two respondents in this age category.

Calculating the median values of the tech-savviness as well as privacy concerns indicates the comparability of both groups in these categories. Both groups reported a median value of three. The range of this question went from *low* to *advanced*, with both groups indicating a median of an *intermediate* tech-savviness score. For privacy concerns related to personal data, the groups both score a median of two, indicating that the respondents were *somewhat concerned* with privacy matters surrounding their personal data.

Gender:			Frequency	Percent	
Least Transparent		Male	19	54,3	
		Female	15	42,9	
		Non-binary / third gender	1	2,9	
		Total	35	100	
Most Transparent		Male	10	37	
		Female	17	63	
		Total	27	100	
Age:			Frequency	Percent	Cumulative Percent
Least Transparent:		18 - 24	11	31,4	31,4
		25 - 34	9	25,7	57,1
		35 - 44	1	2,9	60
		45 - 54	3	8,6	68,6
		55 - 64	10	28,6	97,1
		65 +	1	2,9	100
		Total	35	100	
Most Transparent:		18 - 24	11	40,7	40,7
		25 - 34	2	7,4	48,1
		35 - 44	2	7,4	55,6
		45 - 54	1	3,7	59,3
		55 - 64	9	33,3	92,6
		65 +	2	7,4	100
		Total	27	100	
Tech-Savviness:		N	Minimum	Maximum	Median (Mode)
Least Transparent:		35	2	5	3 (3)
Most Transparent:		27	2	5	3 (3)
Privacy Concerns:					
Least Transparent:		35	1	4	2 (2)
Most Transparent:		27	1	4	2 (2)

Table 7: Descriptives split by Scenario

4.2 Transparency, Trust & Power Asymmetry

The following subchapter will present a comparison of the most and least transparent vignettes by analysing the three newly computed variables: transparency, trust and the power asymmetry.

The dataset was split between the most and least transparent scenario and the median of both transparency items was calculated as well as the interquartile range and the mode. To create a matching interpretation, the combined trust and power asymmetry variables were recoded, inversing the Likert-scale. This was necessary as the statements for both the trust as well as the power asymmetry variables were phrased negatively. Therefore, the interpretation for these variables were that the value five, or *completely agree* on the scale indicated less trust and a widening power asymmetry respectively. By recoding the variables, the new variable indicates less trust and a widening power asymmetry at the value of one.

Transparency, Trust & Power Asymmetry:			TranspC	TrustC	PowerC
Least Transparent	N		35	35	35
	Median (Mode)		1,5 (1)	2 (2)	2 (2)
	Percentiles	25	1	1	1
		50	1,5	2	2
		75	2,5	2	2
Most Transparent	N		27	27	27
	Median (Mode)		2,5 (1,5a)	2 (2)	2 (2)
	Percentiles	25	1,5	1	1
		50	2,5	2	2
		75	3,5	2	2
a Multiple modes exist. The smallest value is shown					

Table 8: Frequencies: Transparency, Trust & Power Asymmetry

Analysing the variables individually and comparing them based on the scenario the respondents were allocated to, it shows that the only difference can be found in the transparency variable. The respondents that were allocated to the least transparent scenario indicated a median value of 1,5 for the effectiveness of transparency while those in the most transparent scenario indicated a median of 2,5. Both sets of respondents showed a value of 2 for the combined variables of trust and power asymmetry.

Following this, I conducted a non-parametric correlation test for the three variables trust, transparency and power asymmetry to search for associations with the coefficient Spearman's Rho.

Spearmans Rho Correlations:					
			Transparency	Trust	Power Asymmetry
Least Transparent	**Transparency**	Correlation Coefficient	1	0,104	-0,095
		Sig. (2-tailed)	-	0,55	0,585
		N	35	35	35
	Trust	Correlation Coefficient	0,104	1	,542**
		Sig. (2-tailed)	0,55	-	0,001
		N	35	35	35
	Power Asymmetry	Correlation Coefficient	-0,095	,542**	1
		Sig. (2-tailed)	0,585	0,001	-
		N	35	35	35
Most Transparent	**Transparency**	Correlation Coefficient	1	0,069	-0,17
		Sig. (2-tailed)	-	0,733	0,396
		N	27	27	27
	Trust	Correlation Coefficient	0,069	1	0,296
		Sig. (2-tailed)	0,733	-	0,134
		N	27	27	27
	Power Asymmetry	Correlation Coefficient	-0,17	0,296	1
		Sig. (2-tailed)	0,396	0,134	-
		N	27	27	27
** Correlation is significant at the 0.01 level (2-tailed)					

Table 9: Correlations between Transparency, Trust & Power Asymmetry

4.2.1 Perception of a widening Power Asymmetry on Trust

The analysis shows that there is only one statistically significant correlation to be found following the Spearman's Rho correlation analysis. Trust and power asymmetry have a statistically significant correlation and a positive correlation coefficient of 0,542 indicating a moderate effect. This allows me to infer that the respondents from the least transparent group with a lower score on the trust variable also perceived a smaller power asymmetry, allowing me to accept H1 for the least transparent scenario:

H1: Citizens that perceive predictive analytics as instruments which widen the power asymmetry between themselves and the government will exhibit a lower score of trust in government.

4.2.2 Effect of Transparency on Trust

The third hypothesis that is presented in Chapter 2.3.4 revolves around how the different scenarios impact the trust that the respondents placed in the hypothetical government. Focussing especially on the most transparent scenario and expecting these respondents to exhibited higher levels of trust, the analysis with the Spearman Rho correlation is statistically insignificant. Although there is a very small positive correlation, it remains statistically insignificant for both the most as well as the least transparent scenario, allowing me to reject H3:

H3: Transparency, in the usage of predictive analytics, positively affects reported trust in government if it is directed upwards as well as inwards and is additionally characterized by the varieties: event & process, real-time and effective transparency.

4.3 Algorithmic Aversion

To assess respondents' attitudes towards more generally how predictive analytics were used in the hypothetical scenario, a statement is presented asking respondents to assess the acceptability of the basis of the decision. Splitting the dataset between the least transparent and most transparent shows that both groups of respondents scored a median value of 4, indicating that they *somewhat agree* that it is unacceptable to base a far-reaching decision such as the halting or termination of welfare benefits on the output of an algorithm. In both groups, *somewhat agree* and *strongly agree* were the most frequent answers with 74,3 % of respondents from the least transparent scenario and 77,7 % of the respondents from the most transparent scenario.

Unacceptability / Algorithmic Aversion:				
		Frequency	Percent	Cum. Percent
Least Transparent	Strongly disagree	1	2,9	2,9
	Somewhat disagree	4	11,4	14,3
	Neither agree nor disagree	4	11,4	25,7
	Somewhat agree	10	28,6	54,3
	Strongly agree	16	45,7	100
	Total	35	100	
	Median (Mode)	4 (5)		
Most Transparent	Strongly disagree	1	3,7	3,7
	Somewhat disagree	3	11,1	14,8
	Neither agree nor disagree	2	7,4	22,2
	Somewhat agree	12	44,4	66,7
	Strongly agree	9	33,3	100
	Total	27	100	
	Median (Mode)	4 (4)		

Table 10: Frequencies Algorithmic Aversion

To further test how the unacceptability of an algorithmic decisions interacts with other variables and to test the second hypothesis, a Spearman's Rho correlation was conducted between the combined transparency variable and the unacceptability / algorithmic aversion variable.

Spearman's Rho Correlations:				
			Transparency	**Unacceptability**
Least Transparent	**Transparency**	Correl. Coeff.	1	-,371*
		Sig. (2-tailed)	-	0,028
		N	35	35
	Unacceptability	Correl. Coeff.	-,371*	1
		Sig. (2-tailed)	0,028	-
		N	35	35
Most Transparent	**Transparency**	Correl. Coeff.	1	0,057
		Sig. (2-tailed)	-	0,777
		N	27	27
	Unacceptability	Correl. Coeff.	0,057	1
		Sig. (2-tailed)	0,777	-
		N	27	27
* Correlation is significant at the 0.05 level (2-tailed).				

Table 11: Correlation between Transparency & Unacceptability

The transparency variable ascends positively, with a value of 5 being the highest transparency score. The unacceptability / algorithmic aversion variable on the other hand, goes in the opposite direction, with a value of 1 indicating that the respondent finds it completely acceptable and 5 completely unacceptable. For the respondents from the least transparent scenario, the Spearman's Rho correlation coefficient is statistically significant. The negative sign indicates that the variables travel in opposite directions. This allows me to infer that the higher a respondent scores on the effectiveness of transparency variable, the more acceptable they also find the usage of a predictive analytics for a far-reaching decision. On the other hand, the association for the respondents in the most transparent scenario is not statistically significant and while the correlation coefficient is positive, it has a very low strength. I must therefore reject the second hypothesis:

H2: Respondents from the least transparent scenario will find the usage of predictive analytics for far-reaching decisions less acceptable than those from the most transparent scenario.

4.4 Perception of Government

In this part of the analysis, I will present the change that respondents indicated in their perception of the government in the hypothetical scenario. First, I will present a table indicating the perceived change of all respondents, before splitting the data set into the most and least transparent scenario.

Perception Change of Government:		Frequency	Percent	Cumulative Percent
Strong negative change		14	22,6	22,6
Negative change		36	58,1	80,6
No change		7	11,3	91,9
Positive change		2	3,2	95,2
Strong positive change		3	4,8	100
Total		62	100	

Table 12: Frequencies: Perception Change of Government

The results show that 80,6% of respondents have a *negative* or *strong negative* change in their perception towards the government. The statements by the respondents that indicated a *strong positive change* did not fit their analysed statement which specified a more negative perception. Similarly, the statement under *positive change* was allocated to a label designated *Unclear*.

4.4.1 Most Transparent and Least Transparent Scenario

Perception Change of Government:			Frequency	Percent	Cumulative Percent
Least Transparent	Strong negative change		8	22,9	22,9
	Negative change		20	57,1	80
	No change		5	14,3	94,3
	Positive change		1	2,9	97,1
	Strong positive change		1	2,9	100
	Total		35	100	
Most Transparent	Strong negative change		6	22,2	22,2
	Negative change		16	59,3	81,5
	No change		2	7,4	88,9
	Positive change		1	3,7	92,6
	Strong positive change		2	7,4	100
	Total		27	100	

Table 13: Perception Change of Government split by Scenario

Splitting the respondents into their respective scenario yields similar results. 80% of the respondents from the least transparent scenario and 81,5% of the respondents of the most transparent scenario indicated that the government's actions changed their perception either *negatively* or *strongly negative*. The following subchapter will qualitatively assess the text box answers of the respondents. By giving respondents the possibility to further clarify their position I can expand the analysis to the specific factors that led the respondents to change their perception.

4.4.2 Qualitative Assessment

In total, 49 respondents clarified their position towards the government and their perception change via the text box question. To analyse these responses, I used the TextIQ tool which is integrated in the Qualtrics programme to give the statements labels and thus group the responses as to why their perception changed. The groups and the number of respondents allocated to each group are indicated in the table below. Some respondents expanded their assessment leading to some responses being allocated to more than one group.

Perception Change of Government:		N	LS	MS
Loss of Trust		19	11	8
Lack of Transparency		11	7	4
Unacceptability of Algorithm		11	4	7
Perceive themselves in the Scenario		4	4	0
Privacy Infringement/Surveillance		3	2	1
Worried of Bias		2	1	1
Aware of Process		1	0	1
Unclear Statement		3	2	1
Total (Total Responses)		54 (49)		

Table 14: Perception Change: Qualitative Grouping

Of the 49 respondents that chose to clarify why their perception changed, 19 respondents were grouped under the Label *Loss of Trust*. All respondents under this group indicated either a *negative* or *strong negative* change of their perception towards the government. Included were proxy statements from respondents such as "The government made a mistake. This would cause me to assume the government will continue to make other mistakes" as well as "Loss of confidence". Statements such as "Being unfairly labelled as a fraud would be very painful. No such decision should be based on an algorithmic model, this system distrusts citizens so citizens would lose trust in government as well [sic]" were allocated to both the label *Loss of Trust* as well as *Unacceptability of Algorithm*.

11 statements were grouped under the label *Lack of Transparency* including 7 from the least transparent group and 4 from the most transparent group. One respondent of the least transparent group indicated "A Government would need to carefully explain what they are doing and why. Any misuse of information must be sanction in a very strict manner in order to create a broad acceptance by the citizens", emphasizing that the explanation is essential. Furthermore, the respondent points out that safeguards should be set in place when using personal information as an input of predictive analytics. A second respondent from the same group indicated "As the usage of data and the analysis of data (drivers of decision) are not transparent – even after the fact", further highlighting that the respondent's negative perception change was driven by a lack of transparency. Under this label, there were also respondents which were allocated under the *Loss of Trust* label such as this respondent: "Fewer trust in government and its actions. The false analysis and especially the little amount of information on the own case lowers trust", who emphasized the clear link that he or she perceives between transparency and trust.

But not only those allocated to the least transparent scenario stated that the lack of transparency was a factor which affected and changed their perception of the government in the hypothetical scenario: "Transparency of data collation [sic] is critical. Any Government's inclusion of citizen opinion or perception is critical. On-line monitoring would be an essential inclusive tool. More research and analysis in this area is essential and a wise move for a Govt body". This respondent highlights the importance of including the opinions and perceptions of citizens when implementing an algorithm and proposes a monitoring approach as a way to citizen inclusion.

The third label is called *Unacceptability of Algorithm* and groups together those respondents whose perception changed due the decision in the hypothetical scenario being based upon predictive analytics. Four respondents from the least transparent scenario indicated that they find the usage of an algorithm unacceptable. One of the respondents from this group stated: "Because I was denied benefits due to an algorithm and not my individual need, which was previously granted based on the paperwork submitted for my personal situation". This respondent stated a strong negative change due to the decision being reversed not based on personal circumstances but through the output of an algorithm. Similarly, to the two labels presented above, there were also overlaps between the labels. A respondent from the most transparent scenario stated: “Simply because concluding someone is guilty based off of an algorithm is inacceptable. Data may be collected and suspects determined [sic] by algorithm, but without further investigation [sic] and definite proof the citizen could not have been proven guilty. The citizen will therefore lose a bit of faith in the government, not necessarily because of the data collecting itself”. Another respondent from this group emphasized the moral hazards he or she perceived in the hypothetical scenario and the usage of algorithms in the citizens case: “Because of the immorality of granting OR denying someone anything at all, based on an algorithm. Regardless [sic] of the accuracy of said algorithm”. Interestingly, this respondent deems it generally unacceptable for algorithmic outputs to be at the basis of decisions which deny or grant something regardless of their accuracy.

The fourth label under which four respondents were grouped is called *Perceive themselves in the Scenario*. All respondents were from the least transparent group and interpreted the question in a manner that they are personally involved thus basing the scenario around their personal circumstances. These responses were not further evaluated.

Three respondents were allocated under the label *Privacy Infringement/Surveillance* and there were no further overlaps with other labels in this group. One respondent stated that the usage of personal data in the scenario brings more disutility forward than utility: “The use of personal data by the government may be in a way that is not benefit citizens, in which it can harm more than the benefit”. A separate respondent voiced concern over the possibility for more governmental control while not making the correct decision in the scenario: “It feels like the government is controlling you and although the government has the capacity (data analysis) they still manage to make the wrong decision”.

Two respondents raised concerns to possible discrimination through the algorithm and were thus grouped under the label *Worried of Bias*. One respondent indicated that the government may discriminate against “honest” citizens and highlighted the importance of auditing and examining the algorithm for biases: “For whatever reason, the algorithm the govt [sic] uses discriminates against folks in my situation who are being honest. It is important for the government to audit and examine their algorithms for unintended bias in the algorithmic functions”. The second respondent from this group proposed that algorithms may have integrated racism: “Not only humans can be racist, but also machines [sic] can be”. While the hypothetical scenario did not clarify the citizens gender, ethnicity or other factors from which respondents could assume racism or discrimination towards the hypothetical individual, this still elevated concern for two respondents.

One respondent from the most transparent scenario was grouped towards the label *Aware of Process*. This respondent stated that putting themselves in the position of the citizen they would believe that the government would clarify the usage of predictive analytics beforehand. Furthermore, the respondent is confident that the indication produced by the algorithm could be clarified and settled if it is a misunderstanding: “Putting myself into the citizen's position – i.e., considering I would had been sufficiently well informed right upfront about “my” governments’ way of handling this type of IT business –, I would have been aware of the fact that all this could possibly happen … and hence would not have been too surprised. In addition, I would be decently confident that this misunderstanding could still be cleared and settled”.

Finally, three statements were either unclear or not related to the question and were thus excluded from a further qualitative analysis.

4.4.3 Transparency and Perception Change

To finalize this chapter and the analysis of the respondents indicated perception change, I used Spearman's Rho to analyse if there is a meaningful statistical association between the combined transparency variable and perception. This allows me to analyse if the judged effectiveness of transparency correlates with the later posed question of how the perception of government would change.

Correlation between Transparency and Perception Change:				
			Transparency	**Perception**
Least Transparent	**Transparency**	Correlation Coeff.	1	,530**
		Sig. (2-tailed)	-	0,001
		N	35	35
	Perception	Correlation Coeff.	,530**	1
		Sig. (2-tailed)	0,001	-
		N	35	35
Most Transparent	**Transparency**	Correlation Coeff.	1	0,201
		Sig. (2-tailed)	-	0,315
		N	27	27
	Perception	Correlation Coeff.	0,201	1
		Sig. (2-tailed)	0,315	-
		N	27	27
** Correlation is significant at the 0.01 level (2-tailed).				

Table 15: Correlation between Transparency & Perception Change

For the respondents in the least transparent scenario, there is a statistically significant correlation between transparency and their change of perception in government. The correlation coefficient of 0,530 points to a moderately strong and positive effect. Due to the positive sign, the variables move in the same direction, allowing me to assume that those respondents in the least transparent scenario with low transparency scores, also scored lower on their future perception of government.

4.5 Perceived Usefulness of Predictive Analytics

This chapter revolves around the second text box question asking respondents give their opinion on the usefulness of predictive analytics. First, I will present the response frequencies of the respondents and thus their sentiment towards the usefulness of predictive analytics. This was the last question of the survey which implies that all respondents had read the hypothetical scenario and answered the follow-up questions, potentially influencing their perception of usefulness. Following the analysis of the multiple-choice question, I will present a qualitative analysis of the text box response in which respondents had the possibility to further clarify why they believe that algorithms such as predictive analytics may be useful for governments.

Taking all respondents into account, 27,4% or 17 respondents indicated that they believe that algorithms such as predictive analytics are either *not at all useful* or only *slightly useful,* with only one respondent opting for the extreme response of *not at all useful*. 31 respondents, or 50% on the other hand believe that these can either be *very useful* or *extremely useful* tools for governments. 14 respondents indicated that they believe that algorithms are *neither* or took a *neutral* stance towards usefulness.

Perceived Usefulness of Algorithms such as Predictive Analytics:				
		Frequency	Percent	Cumulative Percent
All Respondents	Not at all useful	1	1,6	1,6
	Slightly useful	16	25,8	27,4
	Neither/Neutral	14	22,6	50
	Very useful	23	37,1	87,1
	Extremely useful	8	12,9	100
	Total	62	100	
Least Transparent	Slightly useful	10	28,6	28,6
	Neither/Neutral	9	25,7	54,3
	Very useful	11	31,4	85,7
	Extremely useful	5	14,3	100
	Total	35	100	
Most Transparent	Not at all useful	1	3,7	3,7
	Slightly useful	6	22,2	25,9
	Neither/Neutral	5	18,5	44,4
	Very useful	12	44,4	88,9
	Extremely useful	3	11,1	100
	Total	27	100	

Table 16: Frequencies: Perceived Usefulness of Algorithms

Splitting the dataset and analysing the responses dependent on scenario allocation did not produce vast differences in the respective responses. 45,7% of respondents from the least transparent group indicated that algorithms may be *very useful* or *extremely useful* for governments while in the most transparent group 55,5% respondents hold this preference. 25,7% in the least transparent group indicated a *neither/neutral* preference, opposed to 18,5% in the most transparent group.

Usefulness of Algorithms:		N	LS	MS
Depends on Usage		16	10	6
Efficiency Gains		15	7	8
Data for Policy		7	4	3
Error Margin		4	2	2
Unclear		2	2	0
Total		44		

Table 17: Usefulness: Qualitative Grouping

Taking the same approach as in Chapter 4.4, I created distinct labels under which the statements were grouped.

The first label, in which 16 respondents were grouped is called *Depends on Usage*. These respondents varied between *slightly useful* and *very useful* and based their assessment on the domain where algorithms such as predictive analytics are used as well as how the government follows up on the interpretation of outputs such as a fraud indication. One respondent, who believes they can be *very useful* indicated: “The given example can indicate that someone is a fraud, but without data analysis the government would not be able to see a pattern between fraudulent cases. Nevertheless, the final call that somebody is a fraud can't be made on predictive analysis”. This perception was echoed by most respondents in this label, with one respondent again placing emphasis on the relationship of trust between citizens and governments: “How useful as it may be, it should not lead to such decisive action. A fraud [sic] prediction is not proof of fraud, and the relationship between government and citizen should be about trust”.

The second label, counting 15 respondents is called *Efficiency Gains* and groups together those respondents who placed emphasis on enhanced data analysis, the speeding up of processes and investigations and the filtering of information. One such respondent who sees algorithms to be *slightly useful* stated: “Algorithms and predictive analytics are powerful tools - the use of which surpasses the capacity of most government employees”. A further respondent, indicating that they believe algorithms to be *very useful* emphasized that governing a vast amount of people requires such tools to be effective: “Predictive analysis is a crucial and very important issue – and hence an important tool – for governing people and political systems: Due to the sheer size of a population, governing people always requires means and methods of statistics (such as assumptions, forecasts, …) – at least in non-totalitarian systems in which a government depends on not losing public loyalty and followership”.

Under the third label *Data for Policy*, I grouped all statements that highlighted how enhanced information may be beneficial towards policy making and providing public services. “More information is often better than less information. In the previous example regarding welfare, one could argue that the poor outcome was a result of the government having insufficient information”, this respondent believes that more information is beneficial for avoiding poor outcomes such as in the hypothetical scenario. Another respondent who indicated algorithms to be *extremely useful* in their opinion emphasized strategic planning and service enhancement: “Data can be analysed in a way to help the Gov`t to provide better services, plan the future in a more strategic way and stabilize planning forecasts significantly”.

The last topic label created is called *Error Margin* and groups together four respondents, ranging from *not at all useful* to *slightly useful*. The statements under this label highlighted the potentially faulty outputs and the danger of multiplying wrong assumptions. One respondent indicated their belief that a lot of errors could occur which would then lead to anger on behalf of the citizens: “Perhaps it could be useful for the government, but I suppose there will occur a lot of mistakes and consequently the citizen will get very angry and turn away from the government”. Another respondent believes that wrong assumptions could lead to wrong reactions: “If the assumptions are wrong, the reactions could be false and could multiply”.

Finally, two statements were either unclear or unrelated to the question and thus disregarded for the qualitative analysis.

4.6 Discussion

To conclude the Results chapter, I will discuss the previously presented empirical analysis, summarize the hypotheses and finally answer the main research questions of this thesis.

Hypotheses:		Status	Relation
H1	Citizens that perceive predictive analytics as instruments which widen the power asymmetry between themselves and the government will exhibit a lower score of trust in government.	Accepted (for the least transparent scenario)	RQ1
H2	Respondents from the least transparent scenario will find the usage of predictive analytics for far-reaching decisions less acceptable than those from the most transparent scenario.	Rejected	RQ2
H3	Transparency, in the usage of predictive analytics, positively affects reported trust in government if it is directed upwards as well as inwards and is additionally characterized by the varieties: event & process, real-time and effective transparency.	Rejected	RQ2

Table 18: Hypotheses

As presented in the table above, the first hypothesis was accepted, but only for the least transparent scenario and the second and third hypotheses were rejected. The second hypothesis was rejected although the transparency variable and the unacceptability variable move in opposite directions in the least transparent scenario as expected. Nonetheless, the results from the most transparent scenario were not statistically significant which hindered me from drawing further conclusions. The sign was also positive and although the correlation coefficient is very weak, this goes against the hypothesized relationship. The third hypothesis was also rejected as there were no statistically significant correlations to be found. My expectation for this hypothesis was a positive, significant correlation between trust and transparency for the respondents in the most transparent group which did not set in. This would show that the directions and varieties of transparency, as envisioned by Hood and Heald (2006) and transferred to a hypothetical scenario involving predictive analytics, lead to the outcome of trust. Nonetheless, in their research on transparency, Cucciniello et al. (2017) state that the effectiveness of transparency may vary depending on the task at hand (2017: 42). More concretely, the authors state that one form of transparency may not bolster citizens trust, but that does not mean that all forms of transparency are ineffective.

Taking this into account, the following section will further specify and answer the two main research questions that were posed for this thesis.

RQ1: How does the application of predictive analytics affect the relationship of trust between citizens and the state?

The first hypothesis which is related to this research question can only be accepted for the least transparent scenario. This indicates that the citizens who perceived a widening power asymmetry also scored lower on the trust variable. Taking into account some of the key points conveyed in chapter 2.3.1 such as that citizens cannot opt out of government, shows that citizens that perceive the power asymmetry to widen will also have less trust in the government. Analysing the median values of the combined trust variable, respondents from both scenarios scored a value of 2, with 1 being the lowest score. While there can be no baseline of general trust towards a hypothetical government, the question asking respondents for their perception change is able to give further insights into the answering of the research question.

The negative change exhibited by both sets of respondents towards the government is a first indicator of how predictive analytics may more generally impact the relationship between citizens and the government. Expanding on this assessment through a qualitative analysis of the text box question respondents had the possibility to answer indicated that the largest group (19 respondents) indicated that their negative change of perception is due to a loss of trust in government. As presented in chapter 4.4.2, the respondents who indicated that a loss of trust led to their change of perception listed factors such as incompetence, distrusting of the individual citizens, a lack of confidence as well as a disrespect of fundamental values such as "in doubt for the accused".

Deeming the overall negative change in perception exhibited by the respondents as well as the largest group indicating a *loss of trust* as the reason, I conclude that predictive analytics may negatively influence the relationship of trust between citizens and the government. On the other hand, this assessment is limited to the specific boundaries set out in the scenario, the respondents that engaged with the survey and the sector in which it plays out: the welfare state and welfare benefits. A shifting of the power asymmetry is also a factor which plays a relevant role due to the nature of predictive analytics and how citizens may perceive their impact. While the hypothesis that the perception of a widening power asymmetry negatively influences trust only holds for the least transparent scenario, it is nonetheless an indicator of how perception of the technology plays out in the relationship of trust towards the government.

RQ2: Does transparency influence citizens attitudes towards a government that uses predictive analytics?

Both hypotheses, H2 and H3, which were linked to this second research question were rejected, but nonetheless, the data provides for valuable insights. While for H2, I found a statistically significant correlation between transparency and the unacceptability or algorithmic aversion variable in the least transparent scenario, the most transparent scenario yielded no statistically significant results or the expected negative sign. This at least allows me to infer that the respondents from the least transparent scenario that scored low on the transparency variable also later deemed it unacceptable to base decision on an algorithmic output. Furthermore, H3 which tested if the directions and varieties of transparency in the most transparent scenario led to a higher score of trust was also rejected. Nonetheless, the combined transparency variable varied between the scenarios, with a value of 1,5 and 2,5 for the least and most transparent scenario, respectively. This indicates that the perceived median transparency varied between the scenarios in the intended direction that the most transparent scored higher than the least transparent scenario.

On the other hand, the qualitative analysis of why respondent's perception changed also provided insights for this research question. The second largest group indicated that their change was due to a lack of transparency emphasizing amongst other things that there was no opportunity to understand the decision in the scenario, untransparent usage and analysis of the data and a general lack of explanation as to what and why the government is doing what it is doing. To conclude this question, while most statistical tests were insignificant this takes the basis to ardently claim that transparency or a lack thereof influences attitudes towards the implementation and usage of predictive analytics. On the other hand, the text box answers of the respondents indicate the importance of transparency and how it influenced some respondent's perception. But, deeming as respondents from both scenarios indicated a lack of transparency as a reason for their perception change can also lead me to assume that the vignettes did not exhibit the intended differences or that respondents still felt the most transparent scenario to be lacking transparent actions and attributes. I will further clarify this in chapter 5.1, in which I lay out the strengths and limitations of this thesis.

While the statistical tests were not able to confirm the hypotheses H2 and H3, respondents that indicated a perception change and expanded on their assessment chose a lack of transparency as the second largest group. Deeming as neither the scenario nor the follow-up questions mention either trust or transparency, I can conclude that these two factors were highly relevant in the respondent pool. A lack of transparency was relevant for respondents from both scenarios and in some cases as presented in the qualitative analysis in chapter 4.4.2 coincided with a loss of trust.

To conclude the discussion of the results I would like to emphasize the last question, which asked respondents if they believe algorithms such as predictive analytics to be useful. Only one respondent believes that they are of no use for the public sector, while the rest of the respondents were mainly grouped between *slightly useful, neither/neutral* or *very useful*. Important to highlight are the two main labels which were emerged from the qualitative analysis of respondent's assessments of usefulness: *depends on usage* and *efficiency gains*. As highlighted in the theoretical background chapter, the algorithmic governance sphere and its subjects are often immersed in different fields of tension. The assessment by the respondents shows how some respondents will highlight the potential for efficiency gains, while for others the usefulness is directly linked to the intended usage.

5. Conclusion

In this last chapter, I will first present the strength and limitations of this research project and critically assess what elements could be relevant for future research. Following this, the *Society-in-the-Loop* concept by Iyad Rahwan (2017) will be shortly introduced to highlight a potential policy proposal for the future of algorithms such as predictive analytics in the public sector. Finally, I will shortly summarize the key points and findings of this thesis.

5.1 Strengths and Limitations

The first limitation I encountered during the analysis of the results was the relatively low Cronbach's α reliability measure for two of the combined variables, trust and power asymmetry. By adding a pre-test, future research could validate that the variables load on the same factor and can thus be combined. Furthermore, due to the low number of items for the combined variables it was not possible to remove an item and achieve a higher reliability. Thus, more items and potentially more implicit questions on trust and power asymmetry could strengthen future approaches. Furthermore, by conducting a pre-test the perception of the most and least transparent scenario could be investigated further. Due to the hypothetical nature of the scenarios the content and vignettes can be seen as a subjective judgement by the researcher and can resonate differently with the respondents. By pre-testing different scenarios and letting respondents judge which is the most or which is the least transparent would give me a more objective measure of respondent's perspectives and an indication on which vignettes should be refined.

Although I did not ask for respondent's nationalities and elaborated the reasons for this in Chapter 3.1. Future research in this area may find meaningful associations stemming from the respective nationality of respondents. This could have played a role in how respondents assessed the government in the hypothetical scenario even if the only information provided was *a democratic country*. Future studies may either focus on participants from one country or seek to an equal demographic from different countries to uncover if this further influences the variables of interest.

I would also like to lay out the strengths of this study. By adding two open-ended questions, I enabled respondents to share their perspectives on why their perception of the government would change and why they believe algorithms such as predictive analytics to be useful or unuseful tools for governments. Adding this qualitative analysis allowed me to gain insights and explanations from the respondents although some statistical tools lacked power and significance. Grouping together respondent's opinions and perceptions under distinct labels gives insights into interesting paths and inquiries for future research and clearly articulate the concerns and perceptions revolving around algorithms and predictive analytics.

Although the results cannot be generalized to a further extent, the answers given by the respondents show which areas future research interests may extend upon. Especially the group which answered *depends on usage* for the usefulness of algorithms such as predictive analytics may be an insightful avenue to test which usages citizens respond favourably to and which lead to more negative perception changes. Following this statement, the next subchapter will introduce a concept which I believe may help alleviate concerns regarding the usage of algorithms and can be seen as a viable way forward for the technology in the public sector.

5.2 Society-in the Loop

As an outlook as well as policy proposal for algorithms in the public sector, I would like to introduce the *Society-in-the-Loop* concept by Iyad Rahwan (2017).

Rahwan's proposition entails an advancement from the current Human-in-the-Loop system towards a societal consensus in the form of an algorithmic social contract (2017: 1). His concept proposes that society takes place within "the loop" to negotiate the "values of various stakeholders affected […], and monitor […] compliance with the agreement" (2017: 1). These actions and engagements intend to ensure the accountability, fairness and transparency of those algorithms that are used in governing, especially due to the risk of black-box systems which were described in chapter 2.3.3 of this thesis and their opaque nature. His proposition envisions embedding the values of a given society into the "algorithmic governance of societal outcomes that have *broad implications*" (2017: 3; Author's emphasis). Rahwan believes that negative externalities which may follow from algorithmic systems in the governance sphere must be quantified so that trade-offs can be negotiated.

Applying this concept to some answers given by the respondents of the survey, I believe that it can be an integral part of positively moving towards the implementation of algorithms and predictive analytics in the public sector. Deeming as many respondents indicated that their negative perception change of the hypothetical scenario was due to a loss of trust or a lack of transparency, this concept which binds stakeholders into the implementation of the systems may help to alleviate these concerns. Additionally, while many respondents saw at least a slight usefulness of algorithms in the public sector, a large emphasis was placed on how they are used (*Depends on Usage*). Rahwan's concept specifically mentions that taking society's values into account for outcomes with broad implications, which I assume entails a decision such as the termination or halting of welfare benefits that is based on an algorithmic output. Emphasizing the assessments given by the respondents, this intuitive concept has the potential to heed some of the concerns that were presented while also ensuring other core compliance factors such as transparency, thus embedding societal values and boundaries into the technology.

5.3 Summary

In this thesis I set out to explore the features of predictive analytics, its potential and fields of tension in the public sector and its relevance for the relationship of trust between citizens and governments.

Drawing inspiration for my research interest from the SyRI case in the Netherlands, I analyse how the application of predictive analytics may affect the relationship of trust between citizens and the state, considering transparency as a potential intermediary. Building up the theoretical background by laying out the concepts that constitute governance by algorithms, Big Data as well as predictive analytics I present the different fields of tension and contested viewpoints. The second part of the theoretical framework entails the concept of trust and transparency. Beginning with a discussion on a potential widening of the power asymmetry between citizens and their governments, the chapter also introduces the concept of Algorithmic Aversion by Berkeley Dietvorst as well as how transparency and opacity relate to algorithmic applications. Following this, the main theoretical building block, the directions, varieties and outcomes of transparency are introduced. Building upon Hood & Heald (2006), I present a conceptual framework which visualizes the links between the concepts and positions the hypotheses.

Subsequently, the third chapter of this thesis clarifies the research design which was chosen as well as the methodological approach. Two hypothetical scenarios with distinct vignettes were placed within an internet-based survey to collect the data. The vignettes in the scenario were built up from Hood & Heald's work on transparency and the content of the scenario was loosely based on the SyRI case by presenting a welfare beneficiary in a democratic country whose benefits are halted due to the output of a predictive model. Concluding the research design chapter, newly combined variables for trust, transparency and power asymmetry were created and tested for their reliability and normality.

In the fourth chapter, I present the results that were achieved through quantitative as well as qualitative analysis of the surveys. By linking both research questions with different hypotheses, I created more testable and nuanced aspects of the main questions. While the first hypothesis, exploring how trust and the power asymmetry correlate only held for the least transparent scenario, it was nonetheless an indicator what factors may play a role in the trust relationship. The qualitative analysis of the data for the first research question pointed my results to clearer conclusion, that the largest group of respondents indicated that their perception of the government would change due to a loss in trust. While other groups such as the lack of transparency or the general unacceptability of the algorithm also coincide with the theoretical background presented, for the main research question I can conclude that predictive analytics may negatively impact the relationship of trust between citizens and the government.

The second research interest explored how transparency may impact the usage of predictive analytics, specifically if a most transparent case will lead to a higher trust score than a least transparent case. While both hypotheses that were related to this question were rejected, the qualitative analysis of the respondent's perception change also enabled me to gain further insights. Behind loss of trust, a lack of transparency was the second largest label under which the assessments of the text box question were grouped. This indicates that a lack of transparency led to a negative perception change towards the government in the hypothetical scenario. Taking into account that respondents from both sets of groups indicated a lack of transparency also gives me insights into how the respondents judged the scenario. Although the median transparency score varied between both scenarios in the intended direction, the most transparent vignettes created might not have been explicit enough. Finally, drawing from the qualitative analysis of how useful respondents believe algorithms such as predictive analytics to be for the public sector, the reoccurring theme from chapter 2 can be made out, that the two main groups are split between depends on usage and efficiency gains.

Following the results and the discussion, I present the strengths and limitations of the thesis. Focussing mainly on how the survey items and the scenarios may have been improved through pre-testing, the strengths of the study are clearly to be found in the qualitative analysis. This enabled deeper insights into the topics that reduced respondent's perception of the government while also showing how they more generally judge the usefulness of predictive analytics.

As a concept for the future use of algorithms in the public sector, I shorty reiterated on the Society-in-the-Loop proposal by Iyad Rahwan which proposes a social contract for algorithmic governance to counteract some opaque features of the technology and embed societal values to manage and negotiate trade-offs. This concept coincides with some considerations and drawbacks presented by the respondents thus giving an outlook on one of the possibilities to achieve societal compromise on the scope, usage and compliance of and with algorithms in the public sector.

To conclude this thesis, although the survey and results are not representative or generalizable, insights were provided as to how predictive analytics may lead to a loss of trust in the government. Furthermore, transparency plays an important role in the actions of the government as does the rationale behind the usage of algorithms in the public sector.

6. Bibliography

Alessandro, M. et al. (2020): *Transparency and Trust in Government. Evidence from a Survey Experiment*. World Development 138 (2021). Elsevier.

Andres, L. (2012): *Developing Survey Questions*. In Andres (2012): Designing & Doing Survey Research. Sage.

Araujo, T. et al. (2020): *In AI we trust? Perceptions about automated decision-making by artificial intelligence*. AI & Society (2020) 35:611-623. Springer.

Auspurg, K. & Hinz, T. (2006): *Why and When to Use Factorial Survey Methods*. In Auspurg & Hinz (Eds.) (2006): Factorial Survey Experiments. Sage.

Bass, I. (2019): *Administration by Algorithm? Public Management meets Public Sector Machine Learning*. Preprint available under: https://www.researchgate.net/publication/332555598_Administration_by_Algorithm_Public_Management_meets_Public_Sector_Machine_Learning. [Last accessed 19.04.2021].

BCG (2017): *Destination Unknown: Exploring the impact of Artificial Intelligence on Government*. Boston Consulting Group Working Paper. Centre for Public Impact.

Bekker, S. (2021): *Fundamental Rights in Digital Welfare States: The Case of SyRI in the Netherlands*. In: Spijkers O., Werner W.G., Wessel R.A. (eds) Netherlands Yearbook of International Law 2019. Netherlands Yearbook of International Law, vol 50. T.M.C. Asser Press, The Hague

Brown, A. et al. (2019): *Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services*. CHI Conference on Huan Factors in Computing Systems Proceedings (CHI 2019), May 4-9, 2019, Glasgow.

Cambridge Dictionary: *Prediction.* https://dictionary.cambridge.org/de/worterbuch/englisch/prediction. [Last accessed 1.03.2021.

Chawda, V. (2018): *Building trust in government's use of data*. KPMG Insights. https://home.kpmg/xx/en/home/insights/2018/06/building-trust-in-governments-use-of-data.html. [Last accessed 14.02.2021].

Chyung, S.-Y. et al. (2017): *Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale*. Performance Improvement, 56(10), pp.15-23.

Citron, D. & Pasquale, F. (2014): *The Scored Society: Due Process for Automated Predictions*. Washington Law Review, Vol, 89, no. 1, March 2014:1-34. Hein Online.

Cucciniello, M. et al. (2017): *25 Years of Transparency Research: Evidence and Future Directions*. Public Administration Review, Vol. 77, Iss. 1: 32-44. The American Society for Public Administration.

Daniell, K., Morton, A., Rios Insua, D. (2016): *Policy Analysis and Policy Analytics*. Annals of Operations Research, 263:1-13. Springer.

De Rechtsspraak (2020): *SyRI legislation in breach of European Convention on Human Rights*. https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx. [Last accessed 28.01.2021].

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). *Algorithm aversion: People erroneously avoid algorithms after seeing them err*. Journal of Experimental Psychology: General, 144(1), 114–126

European Liberal Forum (2019): *Predictive Analytics and AI in Governance: Data-driven government in a free society – Artificial Intelligence, Big Data and Algorithmic Decision-Making in government from a liberal perspective.* ELF & NEOS Lab.

Gamage, P. (2016): *New development: Leveraging "big data" analytics in the public sector*. Public Money & Management, 36:5, 385-390.

Graham Stone, F. et al. (2016): *Navigating the Transparency – Privacy Paradox in Public Sector Data Sharing*. American Review of Public Administration 2016, Vol. 46(5), pp.569-591. Sage Publishing.

Grimmelikhuijsen, S. (2009): Do transparent government agencies strengthen trust? Information Polity 14 (2009), pp.:173-186. IOS Press.

Grimmelikhuijsen, S. (2012): *Linking transparency, knowledge, and citizen trust in government: an experiment*. International Review of Administrative Sciences 78(1): 50-73. Sage.

Grimmelikhuijsen, S. et al. (2013): *The Effect of Transparency on Trust in Government: A Cross-National Comparative Experiment*. Public Administration Review, July 2013. https://www.researchgate.net/publication/263301206. [Last accessed 29.03.2021].

Grimmelikhuijsen, S., and Welch, E. (2012): *Developing and Testing a Theoretical Framework for Computer-Mediated Transparency of Local Governments*. Public Administration Review 78(1), 562-571. Preprint: https://www.researchgate.net/publication/254886771. [Last accessed 31.03.2021].

Guy Peters, B. & Guedes-Neto, J.V. (2020): *Experimental methods A: survey experiments in public administration*. In Vigoda-Gadot & Vashdi (Eds.) (2020): Handbook of Research Methods in Public Administration, Management and Policy. Edward Elger Publishing.

Henley, J. & Booth, R. (2020): *Welfare surveillance system violates human rights, Dutch court rules*. https://www.theguardian.com/technology/2020/feb/05/welfare-surveillance-system-violates-human-rights-dutch-court-rules. [Last accessed 31.01.2021].

Hinton, P. et al. (2004): *SPSS Explained*. Routledge Publishing.

Höchtl, J., Parycek, P., Schöllhammer, R. (2016): *Big data in the policy cycle: Policy decision making in the digital era*. Journal of Organizational Computing and Electronic Commerce, Vol. 26, NOS.1-2, 147-169. Taylor & Francis.

Jamieson, S. (2005): *Likert Scales: How to (ab) Use Them*. Medical Education 38 (12).

Janssen, M. & Kuk, G. (2016): *The challenges and limits of big data algorithms in technocratic governance*. Government Information Quarterly 33 (2016) 371-377. Elsevier.

Jashari, M. & Pepaj, I. (2018): *The Role of the Principle of Transparency and Accountability in Public Administration*. Acta Universitatis Danubis, Vol. 10, no.1/2018: 60-69.

Katzenbach, C. & Ulbricht, L. (2019). *Algorithmic Governance*. Internet Policy Review, 8(4).

Kurkovsky, S. & Syta, E. (2010): *Digital natives and mobile phones: A survey of practices about privacy and security*. Conference Paper: Technology and Society (ISTAS). https://www.researchgate.net/publication/224158104_Digital_natives_and_mobile_phones_A_survey_of_practices_and_attitudes_about_privacy_and_security [Last accessed 26.04.2021].

Lepri, B., et al. (2017): *Fair, Transparent, and Accountable Algorithmic Decision-making Process: The Premise, the Proposed Solutions, and the Open Challenges*. Philosophy & Technology (2018) 31: 611-627. Springer Science + Business.

Lind, A. (2018): *Transparency, trust and public value*. In Wanna, J., & Vincent, S. (Eds.) (2018). Opening Government: Transparency and Engagement in the Information Age. ANU Press.

Mayer-Schönberger, V. (2015): *Big Data – Eine Revolution, die unser Leben* verändern *wird*. Bundesgesundheitsblatt 2015 – 58:788-793. Springer.

McKinlay, S. (2020): *Trust and Algorithmic Opacity*. In Macnish, K. & Galliott, J. (Eds.), Big Data and Democracy (pp. 153-166). Edinburgh University Press.

Meijer, A. (2013): *Understanding the Complex Dynamics of Transparency*. Public Administration Review, Vol. 73, Iss. 3, pp. 429-439. The American Society for Public Administration.

Meijer, A. et al. (2018): *Assessing Government Transparency: An Interpretive Framework*. Administration & Society, 2018, Vol. 50(4): 501-526. Sage.

Miller, A. & Mitamura, T. (2003): *Are Surveys on Trust Trustworthy?* Social Psychology Quarterly, Mar. 2003, Vol. 66, No. 1, pp. 62-70. American Sociological Association.

O'Reilly, M. (2013): *As Private Sector Embraces Big Data, Public Sector Falls Behind*. https://theglobalobservatory.org/2013/05/as-private-sector-embraces-big-data-public-sector-falls-behind/. [Last accessed 31.01.2021].

Pallant, J. (2016): *SPSS Survival Manual*. Open University Press.

Park, H. & Blenkinsopp, J. (2011): *The roles of transparency and trust in the relationship between corruption and citizen satisfaction*. International Review of Administrative Sciences, 77 (2) pp.: 254-274. Sage Publishing

Pencheva, I. et al. (2020): *Big Data and AI – A transformational shift for government: So, what next for research?* Public Policy and Administration 2020, Vol. 35 (1) 24-44. Sage.

Peters, R. & Schuilenberg, M. (2018): *Machine Justice: Governing security through the bureaucracy of algorithms*. Information Polity 23 (2018) 267-280. IOS Press.

Poel, M. et al. (2018): *Big Data for Policymaking: Great Expectations, but with Limited Progress?* Policy & Internet, 10:3. Policy Studies Organization.

Rahwan, I. (2017): *Society-in-the-Loop: Programming the Algorithmic Social Contract*. Preprint. Ethics of Information Technology. https://arxiv.org/abs/1707.07232. [Last accessed 03.06.2021].

Rodrigues, K.F. (2018): *Unveiling the concept of transparency: its limits, varieties and the creation of a typology*. Cad. EBAPE.BR, v.18, no.2: 237-253. Apr./June 2020.

Taherdoost, H. (2019): *What Is the Best Response Scale for Survey and Questionnaire Design: Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale*. International Journal of Academic Research in Management, 2019, 8.

TechAmerica Foundation (2012): *Demystifying Big Data: A Practical Guide to Transforming The Business of Government*. https://breakinggov.com/documents/demystifying-big-data-a-practical-guide-to-transforming-the-bus/. [Last accessed 21.02.2021]

The Economist (2021): *Spy agencies have high hopes for AI. This isn't their first attempt*. https://www.economist.com/science-and-technology/2021/03/04/spy-agencies-have-high-hopes-for-ai. [Last accessed 07.04.2021].

Tsoukias, A. et al. (2013): *Policy analytics: an agenda for research and practice*. EURO J Decision Process (2013) 1:115-134. Springer.

Van der Voort, H.G. et al. (2019): *Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making?* Government Information Quarterly 36 (2019) 27-38. Elsevier.

Van Schendel, S. (2019): *The Challenges of Risk Profiling Used by Law Enforcement: Examining the Cases of COMPAS and SyRI*. In: Reins L. (eds) Regulating New Technologies in Uncertain Times. Information Technology and Law Series, vol 32. T.M.C. Asser Press, The Hague.

Vydra, S. & Klievink, B. (2019): *Techno-optimism and policy-pessimism in the public sector big data debate*. Government Information Quarterly 36 (2019), 101383. Elsevier.

Waggoner, P. et al. (2019): *Big Data and Trust in Public Policy Automation*. Stat Polit Pol 2019; 10(2): 115-136. De Gruyter.

Waller, M. & Waller, P. (2020): *Why Predictive Algorithms are So Risky for Public Sector Bodies.* London October 2020. Available at SSRN: https://ssrn.com/abstract=3716166.

Wanna, J. (2018): *Opening Government: Transparency and engagement in the information age*. In Wanna, J., & Vincent, S. (Eds.) (2018). Opening Government: Transparency and Engagement in the Information Age. ANU Press.

Universiteit Leiden

Appendix

Informed Consent

Welcome to the research study "Predictive Analytics in the Public Sector".

I am a Master's student at Leiden University, The Hague in the MPA: Economics & Governance programme. I am interested in understanding the usage of Predictive Analytics in the Public Sector and the perception of citizens on this topic. For this study, you will be presented with information relevant to the field of Algorithmic Governance. You will be presented with a hypothetical scenario and then, you will be asked to answer some follow-up questions. Your responses will be kept completely confidential and there will be no analysis of individual responses.

Predictive analytics are an application belonging to the **family of algorithms**. Outputs are generated through computations and statistical analyses. **The goal of making predictions with algorithms is to analyse past data to find meaningful associations which enable a prediction of an event or behaviour in the future**.

The study should take you around 8 to 9 minutes to complete. Your participation in this survey is voluntary. You have the right to withdraw at any point during the study by closing your browser tab. The Principal Investigator of this study can be contacted at s.w.p.scherg@umail.leidenuniv.nl.

By clicking the button below, you acknowledge:

- Your participation in the study is voluntary.
- You are aware that you may choose to terminate your participation at any time for any reason.
- If you do not consent and do not wish to participate in the study, please close this window

If you have further questions regarding the survey, the analysis or the research outcome please feel free to contact me per E-mail.

○ I consent, begin the study

Basic Questions

Please state your gender:

○ Male
○ Female
○ Non-binary / third gender
○ Prefer not to say

Please state your age:

○ 18 - 24
○ 25 - 34
○ 35 - 44
○ 45 - 54
○ 55 - 64
○ 65 +

Please indicate your tech-savviness. To better help you interpret the scale of the answers, please see the table below:

Low	Intermediate	Advanced
Very basic usage of phones, PCs & Laptops but no working ability with programmes or apps.	Usage and skills of an average student or office worker completing basic administrative tasks. Usage of phones, PCs…as well as workability with programmes such as Excel, Word etc.	Describes workability which goes further than standard administrative tasks including programming languages, deep understanding of software and knowledge of complex applications such as algorithms.

- ◯ Low
- ◯ Low/Intermediate
- ◯ Intermediate
- ◯ Advanced/Intermediate
- ◯ Advanced
- ◯ Unsure

Please indicate how concerned you are with privacy matters related to your personal data? Personal data does not only include the data you produce with your smartphone, laptop etc.but also the data you provide to the Government, insurances etc.

- ◯ Very concerned
- ◯ Somewhat concerned
- ◯ Neither concerned nor unconcerned
- ◯ Somewhat unconcerned
- ◯ Not at all concerned

Hypothetical Scenario a.) **[Most transparent scenario; Not visible to Respondents]**

In the following paragraph you will be presented with a hypothetical scenario. Please carefully read the scenario as you will be asked to answer follow-up questions related to the information presented.

A citizen lives in a democratic country and is a welfare beneficiary. The country the citizen lives in publishes widespread, accessible information on governmental activities and actively encourages citizens to scrutinize this information. In an effort to modernize their administrative arms, the Government has decided to utilise predictive analytics to make full use of the data through more efficient analysis and the possibility of predictions. Both where and why this technology will be used is communicated by the Government. After filing all relevant documents, the citizen has been receiving benefits for 6 months. During this time, the payments stop and the citizen receives a letter that the case has been labelled as "potentially fraudulent" which led to the halting of the benefits.
The citizen is familiar with the governmental organization in charge of handling the benefits due to the clear instructions provided by the caseworker at the organization as well as through their quarterly updates. Furthermore, in the letter the citizen received, the organization gives notice that the halting of the decision to discontinue the benefits payment is based upon the output of an algorithm used to analyse data and make predictions. The citizen has filed the application at the relevant office in a timely and orderly manner and is concerned about the discontinuation of benefits. This leads the citizen to contact the organization and inquire why the payments have been terminated.

The answer the citizen receives from the inquiry is an explanation as to which data is collected and used as in input of the algorithm. This led to the case showing similarities to other fraud cases.

○ Proceed to follow-up questions

Hypothetical Scenario b.) **[Least transparent scenario; Not visible to Respondents]**

In the following paragraph you will be presented with a hypothetical scenario. Please carefully read the scenario as you will be asked to answer follow-up questions related to the information presented.

A citizen lives in a democratic country and is a welfare beneficiary. The country the citizen lives in is keen on understanding its citizens and collects and analyses broad swathes of data including with algorithms and advanced analytical applications. After filing all relevant documents, the citizen has been receiving benefits for 6 months. During this time, the payments stop and the citizen receives a letter that the case has been labelled as "potentially fraudulent" which led to the halting of the benefits.
The citizen has had contact with the organization during the application process for the benefits, but due to the little information they publish on their procedures and innerworkings, the citizen feels unfamiliar with the organization. The citizen has filed the application at the relevant office in a timely and orderly manner and is concerned about the discontinuation of benefits. This leads the citizen to contact the organization and inquire why the payments have been terminated.

The answer the citizen receives from his inquiry is that the case has been analysed using a predictive algorithm which assessed the data provided and concluded that there are similarities to past cases of fraud. Trying to gain further insights, the citizen initiates a Freedom of Information appeal hoping to understand the factors that the algorithm considered. The citizen receives a document clarifying that all data provided to the Government may be used as an input to the algorithm. Furthermore, the organization states that they may not publish the exact innerworkings of the algorithm to avoid third parties from "gaming" the system and avoiding detection.

○ Continue to follow-up questions

Follow-Up Questions

Imagine yourself in the citizen's position. Please share your thoughts on the following statement:

The citizen has been sufficiently informed by the Government as to why the benefits have been halted and the case is being investigated

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Do you believe the Government has interacted in a way that the citizen in the scenario has the possibility to understand the decision that was made?

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

Do you believe the citizen will feel uneasy in the future with sharing data, especially data the citizen is obliged to share with the Government?

- Definitely not
- Probably not
- May or may not
- Probably yes
- Definitely yes

Citizens cannot opt out of Government but they can withdraw their support. Do you believe that making predictions on individual citizens behaviour will lead to such a withdrawal?

- ◯ Strongly disagree
- ◯ Somewhat disagree
- ◯ Neither agree nor disagree
- ◯ Somewhat agree
- ◯ Strongly agree

The outcome of the hypothetical scenario and thus if the citizen committed fraud or not is unknown. Please imagine for this question that the citizen <u>did not</u> commit fraud but was labelled as such by the predictive algorithm. Do you believe the citizen will have less faith in the competence of the Government?

- ◯ Definitely not
- ◯ Probably not
- ◯ May or may not
- ◯ Probably yes
- ◯ Definitely yes

It is unacceptable for the Government to base a decision which has such far-reaching consequences as the halting or termination of welfare benefits on the output of an algorithm.

- ◯ Strongly disagree
- ◯ Somewhat disagree
- ◯ Neither agree nor disagree
- ◯ Somewhat agree
- ◯ Strongly agree

Why does the citizen take action in the hypothetical scenario?

- ○ Bail is not granted
- ○ Rejection of an asylum application
- ○ Application to a mortgage credit is declined
- ○ Welfare benefits are halted

Please share your thoughts on the following statement:

The Governments usage of predictive analytics enables greater surveillance of citizens even if this is not intended

- ○ Strongly disagree
- ○ Somewhat disagree
- ○ Neither agree nor disagree
- ○ Somewhat agree
- ○ Strongly agree

Information on individual citizens becomes more visible for the Government.

- ○ Strongly disagree
- ○ Somewhat disagree
- ○ Neither agree nor disagree
- ○ Somewhat agree
- ○ Strongly agree

The possibility to link and analyse data in such a manner maylead to privacy infringements of individuals.

- ○ Strongly disagree
- ○ Somewhat disagree
- ○ Neither agree nor disagree
- ○ Somewhat agree
- ○ Strongly agree

If you were in the citizen's position, would your future perception of the Government change?

- ◯ Strong negative change
- ◯ Negative change
- ◯ No change
- ◯ Positive change
- ◯ Strong positive change

Please shortly indicate why your perception of the Government would change? If you answered *No change* on the previous question you may leave the box empty.

Can algorithms such as predictive analytics be a useful tool for Governments in your opinion?

- ◯ Not at all useful
- ◯ Slightly useful
- ◯ Neither/Neutral
- ◯ Very useful
- ◯ Extremely useful

Please briefly explain your assessment. If you answered Neither/Neutral, you may leave the box empty.