Can explainable ai mitigate decision-making errors induced by algorithms in street-level police work? An experiment.

submitted in partial fulfillment for the degree of master of science

Friso Selten
5530709

research master in public administration and organisational science
faculty of law, economics and governance
utrecht university

2021-07-14

	First Supervisor	Second Supervisor	Practical Supervisor
Title, Name	Dr. Stephan Grimmelikhuijsen	Prof. dr. Albert Meijer	Marcel Robeer MSc
Affiliation	Utrecht University	Utrecht University	Dutch Police
Email	s.g.grimmelikhuijsen@uu.nl	a.j.meijer@uu.nl	marcel.robeer@politie.nl

Universiteit Utrecht

PREFACE

In this thesis, I researched how algorithms influence street-level decision-making. This thesis marks the completion of my Research Master's program in Public Administration and Organisational Sciences and was written as part of an internship at the Dutch Police.

I would like to thank several people who helped me conduct this research. Bram van der Linden, Tjitte Westrik, Joost Kroese, and Wout Bouve for their help with distributing the survey. Additionally, I would like to thank everyone who commented on draft versions of the experimental design. In particular, Wout Boeve and Susan van der Klauw. Furthermore, I am grateful to all the anonymous police officers who completed the survey.

I also want to thank my three supervisors. Stephan Grimmelikhuijsen for being an inspiring and enthusiastic first supervisor. Stephan's feedback not only strengthened this research but also helped me advance as a researcher. Marcel Robeer supervised this research from the Dutch Police. Marcel provided crucial insight into the field-specific side of the use of algorithms within the police. As a second supervisor, Albert Meijer provided valuable feedback on the thesis proposal that helped strengthen both the methodological and theoretical aspects of this thesis.

This thesis has was written in the form of an extensive article. I have tried to be concise where possible and detailed where necessary. I hope it provides an insightful and interesting read.

Friso Selten
Utrecht, July 14, 2021

TABLE OF CONTENTS

Can Explainable AI Mitigate Decision-Making Errors Induced by Algorithms in Street-Level Police Work? An Experiment.

Friso Selten, 5530709, f.j.selten@uu.nl
Research in Public Administration and Organisational Science, Utrecht University

ABSTRACT

Machine learning algorithms are increasingly used in the street-level bureaucracy. Frontline decision-making at the same time demands individual and human judgment that cannot be fully automated. Algorithms are therefore used to inform, but not replace the street-level bureaucrat. Street-level decision-makers can however become subject to automation bias or confirmation bias when interpreting algorithmic information. This respectively means that decision-makers overly or selectively trust algorithmic advice. These biases can lead to new types of decision-making errors. Explainable Artificial Intelligence techniques, algorithmic systems that explain how advice is constructed, are seen as a critical step in preventing algorithm-induced decision-making errors. A pre-registered survey experiment was used to test these expectations in a mock algorithm, within a sample of street-level police officers (N = 124). The results of this experiment imply that (1) street-level bureaucrats are *not* prone to automation bias, rather (2) they *are* likely to be subject to confirmation bias. Additionally, this study finds that (3) the effects of explaining algorithmic advice might only be *limited* for professional decision-makers. These findings have important implications for how street-level decision-making processes can be enabled by algorithms.

1. INTRODUCTION

National and international regulators repeatedly emphasize the importance of human oversight over machine learning algorithms (Wagner; 2016; Jung, Mueller, Pedemonte, Plances & Thew, 2019; European Commission, 2021). Human control over algorithmic systems is required to provide a fair assessment of individual cases and to prevent algorithmic mistakes (Binns, 2020). From a public administration perspective, studying machine learning is therefore becoming increasingly important. Machine learning approaches automatically identify patterns between input data and outcomes in large amounts of training data, which they use to predict outcomes on as-yet-unseen data (Young, Bullock & Lecy, 2019). These approaches enable the automation of increasingly complex tasks; machine learning algorithms are starting to enhance and advise in frontline decision-making (Young et al., 2019; Bullock, 2019; Zouridis, van Eck & Bovens, 2020). As the human-in-the-loop, street-level bureaucrats are therefore increasingly asked to control algorithmic functioning (Busuioc, 2020). Controlling algorithmic functioning can however be difficult (Bainbridge, 1983; Zerilli, Knott, Maclaurin & Gavaghan, 2019; Peeters, 2020; Bayamlıoğlu, 2021).

This study investigates two biases in human decision-making that fundamentally hinder the controllability of algorithmic systems. The first is *automation bias*. This bias describes that decision-makers can become overconfident in the rationality of machine learning systems (Lyell & Coiera 2017; Alon-Barkat & Busuioc, 2021). Research in aviation and public health demonstrated that decision-makers are susceptible to automation bias (Lyell & Coiera, 2017). At the same time, there is evidence that street-level bureaucrats are better able to contest automated advice (Weller, 2006; Keddell, 2019). Alon-Barkat and Busuioc (2021) show that, rather than being subject to automation bias, street-level bureaucrats might be susceptible to *confirmation bias* when interpreting algorithmic outputs. This is the tendency of decision-makers to search for evidence that confirms their prior beliefs, while contradictory evidence is neglected (Klayman, 1995; Jones & Sugen, 2001).

Automation bias and confirmation bias can lead to new decision-making errors. If decision-makers automatically conform to algorithmic advice even when faced with contradictory evidence, the human-in-the-loop will not prevent algorithmic mistakes (Skitka, Mosier & Burdick, 1999). A certain degree of confirmation bias, therefore, prevents the occurrence of automation bias. Confirmation bias entails that decision-makers weigh algorithmic predictions to their prior beliefs and only selectively follow algorithmic recommendations that are congruent with these beliefs. Confirmation bias, however, can have negative effects. Street-level bureaucrats will not be able to correct prejudicial algorithmic advice that matches their

pre-existing stereotypical beliefs. (Kassin, Dror, & Kukucka, 2013; Alon-Barkat & Busuioc, 2021). The first aim of this thesis is to investigate the extent to which algorithms induce automation bias or confirmation bias in street-level decision-making processes.

The negative effects of machine learning-induced biases are, at least to some extent, expected to be caused by a lack of understanding of these technologies (Burrell, 2016; Peeters, 2020). Explaining algorithmic functioning is therefore, both in academia and in practice, described as being one of the fundamental elements to prevent decision-making biases induced by automation (Rader, Cotter & Cho, 2016; Ananny & Crawford, 2018; Bannister & Connolly, 2020; European Commission, 2021). Consequently, there is a growing demand for machine learning approaches that not only perform well, but that are also transparent, interpretable, and trustworthy (Holzinger, Mak, Kieseberg & Holzinger, 2018; Giest & Grimmelikhuijsen, 2020). This is the goal of a specific area of machine learning research called explainable AI (XAI); developing algorithms that make machine learning approaches understandable (Adadi & Berrada, 2018).

There is, however, little empirical evidence on the effects of XAI in complex frontline public decision-making processes (Peeters, 2020; Giest & Grimmelikhuijsen, 2020). Research into XAI demonstrated that comprehensibility of algorithmic systems impacts decision-making speed, accuracy, and confidence (Huysmans, Dejaeger, Mues, Vanthienen & Baesens, 2011), and that explanations enable spotting algorithmic mistakes (Ribeiro Singh & Guestrin, 2016). At the same time, XAI can have negative effects. Van der Waa, Nieuwburg, Cremers, & Neerincx, (2021) demonstrated that explanations can persuade users to follow incorrect advice. Overall, the empirical knowledge on the impact of XAI is limited, specifically within complex public decision-making processes (Adadi & Berrada, 2018). The second aim of this thesis is to investigate the effect that explaining algorithmic functioning has on how street-level bureaucrats use algorithmic advice.

These two central aims – studying by algorithm-induced decision-making biases and the effects of XAI – are investigated for a classic street-level bureaucrat: the police officer (Lipsky, 1980; Maynard-Moody & Musheno, 2003). Insights into the influence that algorithms have on policing are especially relevant as this is one of the largest public-sector areas in which algorithms are being implemented (Bullock, Young & Wang; 2020; Peeters, 2020; Meijer, Lorenz & Wessels, 2021). The Dutch police is at the forefront of this adoption. The organization uses machine learning to forecast high crime risk areas, to pre-identify young offenders, to analyze vehicle movement patterns, and to assist citizens with crime reporting (Dechesne et al., 2019; Meijer & Wessels, 2019; Rathenau Instituut, 2019). Given the explicit focus on automation bias and confirmation bias, it is specifically investigated how police officers utilize algorithmic advice that is congruent and incongruent with their professional judgment. This is researched by answering the following question: *Does algorithmic advice introduce automation bias and confirmation bias in street-level decision-making, and can explainable AI help to mitigate the negative effects of these biases?*

This question is answered using a survey experiment conducted within a population-based sample of street-level police officers (N = 124). A realistic and in depth-understanding of the central concepts is provided by testing the effects of a mock algorithm that assists police officers with fencing off the area of a crime. This algorithm is based on an algorithm currently being developed by the Dutch police. Additionally, effects are tested in three highly mundane experimental scenarios: a burglary, an ATM robbery, and a stabbing incident, to account for contextual differences. The findings of this study have important implications both for theory and for practice.

Specifically, the contribution of this thesis is threefold. The research findings first contribute to public administration literature on the use of algorithms within the public sector. In public administration, two distinct views on the use of algorithms exist (Buffat, 2015; Busch & Henriksen, 2018). A *curtailing perspective*, that links machine learning to an increase of unfairness in public decision-making (Peeters, 2020; van Eijk, 2020), and an *enabling perspective*, which sees machine learning approaches as a tool to improve decision making quality (Brundage et al., 2018; Bullock, 2019; Binns, 2020). In this thesis, I researched these two views by focusing on decision-making biases that algorithms possibly induce. The results of this study indicate police officers are more prone to confirmation bias than to automation bias when interpreting algorithmic outputs. This mostly supports an enabling view on the use of algorithms in the street-level bureaucracy but also highlights new risks. This finding adds knowledge to the emerging body of behavioral

public administration literature that investigates biases in decision-making in general, and induced by automation specifically (Battaglio, Belardinelli, Bellé & Cantarelli, 2019; Alon-Barkat & Busuioc, 2021; de Boer & Raaphorst, 2021). In doing so, this research provides initial answers to the numerous calls for more empirical insight into (1) the influence algorithms have on street-level bureaucrats (Bullock et al., 2020; Gritsenko & Wood, 2020; Busuioc, 2020), and (2) the effects of algorithmic transparency and explainability in a public administration context (Peeters, 2020; Giest & Grimmelikhuijsen, 2020).

The second contribution is to XAI literature. Explaining algorithmic functioning is often seen as a critical step to increase trust in algorithmic systems and to prevent algorithm-induced decision-making mistakes (Holzinger et al., 2018; Zerilli et al., 2019; Weller, 2019). Empirical evidence for the effects of explaining algorithmic functioning is, however, limited and inconclusive (Adadi & Berrada, 2018). The findings in this study are sobering. They imply that providing explanations only has a very limited effect on how street-level bureaucrats use algorithmic recommendations. Solely explaining algorithmic functioning can therefore not be expected to enable the correct use of algorithmic advice.

The third contribution is to practice. The findings indicate that frontline workers do not automatically conform to algorithmic advice, but weigh this to their prior knowledge and beliefs. This shows that the implementation of algorithms in frontline work does not automatically lead to an increase in unfair decision-making. Moreover, it highlights that algorithmic systems need to be designed to direct decision-makers towards desired behavior, but simultaneously leave room for adaptability and flexibility. These are important insights for public organizations and regulators that assign a human-in-the-loop to prevent decision-making errors induced by algorithms.

2. ALGORITHMS IN THE STREET-LEVEL BUREACRACY

Algorithms have the potential to fundamentally alter the work of street-level bureaucrats and how public services are delivered (Bullock, 2019; Giest & Grimmelikhuijsen, 2020). In this section, the influence machine learning algorithms have on street-level decision-making are explored. This highlights that algorithms can enable or curtail frontline decision-making quality. XAI might be crucial to support the enabling function of algorithms. Four hypotheses are formulated to test these expectations.

2.1. Administrative and Algorithmic Discretion

Street-level decision-making is characterized by the exercise of administrative discretion (Maynard-Moody & Musheno 2003, p. 9). To comprehend the impact of algorithms on street-level decision-making, it is therefore necessary to understand the nature of administrative discretion itself. Administrative discretion is caused by a mismatch between general rules and local situations. Public officials are expected to base their decisions on pre-defined laws, procedures, and standards (Davis 1969). These general rules and regulations, however, do frequently not correspond to the complex local realities of frontline work. Street-level bureaucrats translate general rules and competing values into client-level decisions (Tummers & Bekkers, 2014). According to Lipsky (1980) this constitutes administrative discretion: "the freedom that street-level bureaucrats have in determining the sort, quantity and quality of sanctions and rewards during policy implementation" (c.f. Tummers & Bekkers, 2014, p. 529).

Administrative discretion has positive and negative consequences. The advantage of administrative freedom is that it allows for experience, local knowledge, sympathy, empathy, insight, and flexibility to be included in frontline work (Maynard-Moody & Musheno 2003; Bullock, 2019; Bannister & Connolly, 2020). Administrative discretion allows decisions to be targeted to the specifics of the local situation. These discretionary practices are not always desirable. The translation of general rules to local decisions is grounded in imperfect information, expertise, and the views of fairness and appropriate action of the street-level bureaucrat itself (Maynard-Moody & Musheno 2003; Bannink; 2018). Additionally, human decision-making is bounded by cognitive limitations (Simon, 1947; Kahneman, 2011). Consequently, administrative discretion has been linked to reducing policy-making effectiveness and efficiency, biased and discriminatory decision-making processes, and unlawful and corruptive behavior (Young et al., 2019; Binns, 2020; Pierson et al., 2020). While allowing for the creation of tailor-made solutions, the negative outcomes of administrative discretion imply it should also be controlled (Davis 1969).

Machine learning algorithms are a new mechanism for exercising control over administrative discretion. Algorithmic approaches quickly analyze available information and apply the same rules to similar situations. Rather than making individualized decisions, machine-learned rules infer based on the extent to which an individual case shares characteristics with a group of other cases (Hannah-Moffat, 2013; van Eijk, 2020). Algorithms therefore can enhance public decision-making accuracy, consistency, objectivity, and efficiency (Le Sueur 2015; Brundage et al., 2018; Bullock, 2019; Binns, 2020). In the street-level bureaucracy, algorithms are for this reason implemented to enhance the judgment of the individual street-level bureaucrat (Meijer et al., 2021). This process, where algorithms are used to augment or automate the exercise of administrative discretion, is what Young et al. (2019) refer to as 'algorithmic discretion'.

While controlling administrative discretion, algorithmic discretion in most frontline tasks does not completely replace the need for human judgment. In fact, the elements that make algorithmic discretion strong – providing efficient, generalizable, and consistent judgment – are simultaneously a weakness. Public decision-making demands case-by-case judgment based on local information (Binns, 2020). Machine learning approaches are not able to deliver such individual judgment and therefore limit an organization's ability to provide a fair assessment over individual cases (Lipsky, 1980; Bannister & Connolly, 2020). In sum, human control over complex and uncertain discretionary tasks is essential to (1) assure individual administrative justice (Binns, 2020), (2) adapt decisions to specific circumstances (Peeters & Widlak, 2018), and (3) override algorithmic errors (Peeters, 2020; Giest & Grimmelikhuijsen, 2020).

These three functions of human control are the motive for implementing algorithms as 'decision-support systems' rather than as autonomous agents (Veale & Brass, 2019). In decision-support systems, algorithms inform and augment decision-making, but a human decision-maker is kept 'in-the-loop' to control algorithmic outcomes (Yeung 2018; Bullock, 2019; Busuioc, 2020). Here, the algorithm offers additional and generalizable information that enhances frontline work, but the street-level bureaucrat in-the-loop ensures individual judgment is provided (Busch & Henriksen, 2018). The expectation in decision-support systems is that, by altering the choice architecture of the decision-maker, algorithmic advice 'nudges' the street-level bureaucrat towards 'desired behavior', while simultaneously leaving room for administrative discretion (De Boer & Raaphorst, 2021). This complementary perspective, where administrative discretion and algorithmic discretion are mutually reinforcing, is what Buffat (2015) refers to as the '*enablement thesis*'.

Table 1. Central concepts and definitions

Concept	Definition
Decision-support system	Algorithmic systems that enhance and advice, but that do not replace the human-decision maker.
Administrative discretion	The freedom that street-level bureaucrats have in determining the sort, quantity and quality of sanctions and rewards during policy implementation.
Algorithmic Discretion	The replacement of administrative discretion by an algorithm that augments or automates a public decision-making task.
Enablement Thesis	The positive expectation that a synergy between administrative and algorithmic discretion improves frontline decision-making quality.
Curtailment Thesis	The negative expectation that algorithmic discretion overly restricts administrative discretion and reduces frontline decision-making quality
Algorithmic Trustworthiness	The willingness to alter decisions based on algorithmic advice and attitudes towards this advice.
Automation Bias	The tendency to use automated cues as a heuristic replacement for a decision-maker's own professional judgment.
Confirmation Bias	The tendency to only incorporate algorithmic advice that fits prior beliefs, while disregarding contradictory evidence.
Explainable AI (XAI)	Algorithmic systems that provide users with an understanding of how the algorithm constructs its predictions.

The enablement thesis presents an optimistic view; algorithms however can also curtail street-level decision making (Buffat, 2015). This '*curtailment thesis*' describes that algorithms overly restrict the exercise of

frontline discretion (Busch & Henriksen, 2018). A fundamental condition under which machine learning algorithms curtail is when they overly reduce human control. This is not an unlikely scenario. Oversight over machine learning approaches is complicated given the large amounts of data they analyze and the complex pattern analyses they perform (Burrel, 2016). Decision-makers therefore often do not have the knowledge, mental capacity, or time to critically assess the functioning of an algorithm. Automation of decision-making can therefore result in the occurrence of new decision-making biases (Peeters, 2020). A human-in-the-loop is therefore no guarantee to correcting algorithmic errors (van Eijk, 2020).

In the remainder of this Section, I formulate two hypotheses to test for the occurrence of two biases that algorithms might induce: automation bias and confirmation bias. Explaining how algorithms produce predictions is often seen as a method to overcome the negative effects of these biases. I specify two additional hypotheses to test this expectation. Table 1 presents an overview of the central concepts used in this study.

2.2. Automation Bias and Confirmation Bias

A central theme in public administration is how decision-making processes work, and how these are influenced by the cognitive limitations of decision-makers (Simon, 1948; Kahneman, 2011; Battaglio et al., 2019). However, only recently empirical work started to investigate how algorithms affect public decision-making (Alon-Barkat & Busuioc, 2021; de Boer & Raaphorst, 2021). In this study, I focus on two fundamental biases algorithms might induce in frontline decision-making.

The first bias central in this study is automation bias. This is "the use of automated cues as a heuristic replacement for vigilant information seeking and processing" (Mosier, Skitka, Heers & Burdick, 1998, p. 48). Automation bias can lead to new decision-making errors, especially when decision-makers overly rely on 'less-than-perfect automation' (Lyell & Coiera 2017). Two types of decision-making errors can emerge: errors of omission and errors of commission. Errors of omission occur when decision-makers fail to take appropriate action because the algorithm does not provide a warning, even when other indicators signal that action is required. Errors of commission occur when decision-makers follow an algorithmic prediction, even in the face of more valid or reliable indicators that suggest that the advice is wrong (Skitka et al., 1999). The use of algorithms then also might not reduce decision-making mistakes, but only lead to a new type of decision-making errors (Skitka, et al., 1999).

Automation bias has been extensively documented in other sectors that became automated, but whether this will also happen in the street-level bureaucracy is unclear. In aviation and health, for example, very reliable algorithms are mostly assisting in simple deterministic decision-making tasks such as monitoring and diagnosis (Lyell & Coiera, 2017). Frontline work is more complex and uncertain (Young et al., 2019). There are indications that decision-makers within these complex public decision-making tasks are able to resist algorithmic advice (Peeters, 2020; Binns, 2020). In an ethnographic study, Weller (2006) demonstrated that street-level bureaucrats are able to use local knowledge to correct automated advice. Likewise, child protection professionals displayed reservations regarding working with predictive algorithms they did not understand (Keddell, 2019). Similar evidence was found in an experimental study by Alon-Barkat and Busuioc (2021). These authors experimentally demonstrated that policymakers, when provided with expert and algorithmic advice, are more likely to follow the advice of the human expert. These findings highlight that automation bias might be less prominent in the uncertainty of the street-level bureaucracy.

This also leads to the second bias that is central in this research: confirmation bias. Alon-Barkat and Busuioc (2021) show that policymakers will not conform to algorithmic advice in general, rather they only rely on algorithmic advice when it matches their pre-existing beliefs and stereotypes. This indicates that public sector workers are not subject to automation bias but to confirmation bias. This bias describes that decision-makers incorporate information that fits their prior beliefs in their decision-making while disregarding contradictory evidence (Nickerson, 1998; Klayman, 1995). The occurrence of confirmation bias in human decision-making is widely documented (Jones & Sugden, 2001; Tschan et al., 2009). Instead of an automatic conformation of algorithmic recommendations, in complex decision-making processes complying with algorithmic advice might thus only occur selectively.

Confirmation bias can have adverse effects, but a certain degree of confirmation bias warrants the controllability of algorithms. The negative consequence of confirmation bias is that street-level bureaucrats will not be able to correct prejudicial algorithmic advice that matches their pre-existing stereotypical beliefs. This is especially concerning because machine learning algorithms are prone to reproduce the prejudiced biases present in human decision-making (O'Neil, 2016; Diakopoulos, 2016). Confirmation bias can therefore lead to algorithms increasing unfairness of street-level decision-making processes (Kassin et al., 2013; Alon-Barkat & Busuioc, 2021). At the same time, confirmation bias to some extent is required for the proper use of predictive systems. Blind trust in algorithmic advice, causing automation bias, can induce new decision-making errors (Skitka et al., 1999). In situations of complexity and uncertainty, decision-makers should therefore not alter their professional judgment based only on one contradictory algorithmic cue (c.f. Klayman, 1995). Confirmation bias from this perspective preserves the values of administrative discretion.

By investigating the occurrence of automation bias and confirmation bias in street-level decision-making, this study assesses whether algorithms can be enabling or if they are mostly constraining frontline work. The occurrence of these biases can however not be directly measured in complex decision-making procedures. Algorithmic systems that infer in complex discretionary street-level tasks are mostly predictive; they advise on tasks without a pre-defined optimal outcome (Young et al., 2019; Bullock et al., 2020). In turn, street-level bureaucrats cannot decide to use automated advice because they know it is correct, rather they need to make this decision based on whether they trust or distrust the advice. To obtain insights into the risks of the two central biases, this thesis focuses on what in the next section is defined to be *algorithmic trustworthiness*.

2.3. Linking Decision-making Biases and Algorithmic Trustworthiness

To conceptualize algorithmic trustworthiness, first, a general definition of trust is essential. A disciplinary transcending definition, that is also commonly used within public administration, is provided by Rousseau, Sitkin, Burt & Camerer (1998, p. 395). These authors define trust to be "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another". This definition indicates that trust consists of two distinct aspects: positive expectations regarding the attitude and behavior of the object of trust (Kramer & Lewicki 2010; Grimmelikhuijsen, Porumbescu, Hong & Im, 2013). Algorithmic trustworthiness is thus both: (1) a willingness of the decision-makers to alter their decision based on algorithmic advice – *trusting behavior*, and (2) a positive attitude of the decision-maker towards this advice – *perceived trustworthiness*. High algorithmic trustworthiness of automated advice can lead to both automation bias and confirmation bias, however in different situations.

Automation bias occurs when the algorithmic trustworthiness of automated advice by the decision-maker is in general high. In human-in-the-loop decision-making systems, street-level bureaucrats are expected to use algorithmic advice, but also to weigh this advice to their own reasoning. If algorithmic trustworthiness of advice is high, the algorithmic prediction will always prevail even when this is incongruent with the professional judgement of the street-level bureaucrat. Essentially, this leads to the human decision-maker being in-the-loop on paper, while in practice no control is exercised (Binns, 2020; van Eijk, 2020).

Confirmation bias occurs when the algorithmic trustworthiness of automated advice by the decision-maker is selective. Following confirmation bias theory, algorithmic trustworthiness will be high when advice is congruent with the professional judgment of the street-level bureaucrat, but low when advice is incongruent with prior beliefs. As described above, this has positive and negative implications. Decision-makers must always remain critical of algorithmic advice, even when this is congruent with their own beliefs. However, this type of confirmation bias, where the street-level bureaucrats use their own professional judgment to value algorithmic outputs, can concurrently be regarded as the very function of the human-in-the-loop of a decision-support system (Young et al., 2019; Veale & Brass, 2019).

This study investigates the occurrence of automation bias and confirmation bias by assessing the trust of police officers in algorithmic advice. Hypothesis 1.1 tests the effect on *trusting behavior*. This is tested by comparing the decisions police officers make when choosing based on their professional judgment, compared to making the same choice when provided with algorithmic advice. This is tested both for algorithmic advice that is congruent and incongruent with the professional judgment of the police officers.

I expect police officers are subject to automation bias and confirmation bias. It is therefore hypothesized that they are likely to follow both the congruent and incongruent algorithmic advice. To gain a more in-depth understanding of whether the effects of automation or confirmation bias are strongest, Hypothesis 1.2 focuses on the second aspect of *algorithmic trustworthiness*. Here, a comparison is made between the perceived trustworthiness of advice that is congruent and incongruent with the professional judgment of police officers.

H1.1) Police officers are more likely to choose an option when this is advised by an algorithm compared to making the same choice based solely on their professional judgement.

H1.2) Police officers perceive algorithmic advice that is congruent with their professional judgment as more trustworthy than algorithmic advice that is incongruent with their professional judgment.

2.4. The Effect of Explainable AI on Decision-Making Biases

Controlling machine learning output is difficult because these technologies are often not understandable for the decision-maker (Burrel, 2016; Peeters, 2020). Explaining how algorithmic advice is constructed is seen as a fundamental element to preventing the negative consequences of biases in decision-making induced by machine learning (Ribeiro et al., 2016; Ahmad, Teredesai and Eckert, 2018).

Explanations are expected to increase the understandability of the user of algorithmic systems. Explanations give decision-makers insight into how advice is constructed and therefore enable informed decision-making. Not all algorithmic explanations, however, provide understanding to a non-technical audience such as street-level bureaucrats. Specifically, Doran, Schulz and Besold (2017) distinguish between opaque systems, interpretable systems, comprehensible systems, and explainable systems. The first system type, a completely opaque system, does not correspond any information about the functioning of the algorithm to the user. These systems solely provide advice without any explanation or extra information. Interpretable and comprehensible systems inform users on which data was used to construct a prediction. However, these systems merely display – respectively in mathematical terms or human-understandable symbols – the information that was used. What these two systems do not show is *why* this data was used. According to Doran et al. providing understanding to non-technical audiences also includes answering this *why-question*. Truly explainable algorithms that support frontline work should thus not only display which data was used but also provide a line of reasoning about why the data was used.

Two approaches can be distinguished in XAI literature to answer this why-question: *global explanations* that explain algorithmic procedures in general, or *local explanations* that explain outcomes in specific situations (Kizilcec, 2016; Lundberg & Lee, 2017; Ahmad et al., 2018, Adadi & Berrada, 2018). Global explanations explain the functioning of the algorithmic system; they explain the general procedures of the algorithm. Global explanations do not show how specific advice was constructed. Explanations that do provide advice-specific insights are referred to as local explanations (Ribeiro et al., 2016). The discretionary tasks of street-level bureaucrats involve case-by-case judgment. Local explanations are therefore most suited to support frontline work. This research therefore explicitly focuses on the effect of explanations that answer why-questions in local situations.

Given the public administration perspective of this thesis, I draw on insights from social science to construct the local explanations rather than on XAI literature. In particular, the work of Miller (2019) forms the basis to construct the explanations. In his article, Miller combines insights from the social sciences and communication sciences to provide guidance on how algorithmic functioning can best be explained. Consequently, although what is technologically possible is taken into account, the construction of the explanations is primarily inspired by the literature that describes how to effectively convey information to a non-technical audience.

Specifically, explanations are formulated that are: everyday, contrastive, and simple (Miller, 2019). Everyday explanations provide explicit cues about why an algorithm constructed specific advice (Lipton, 1990). Contrastive explanations demonstrate why a specific prediction was advised in comparison to a prediction that was not advised (Mercado, et al., 2016). Furthermore, the explanations are kept simple. Simpler explanations have been demonstrated to be more effective in communicating information to users

(Thagard, 1989; Read & Marcus-Newhall, 1993). The work of Miller and the XAI literature suggests that these types of everyday, contrastive, and simple explanations are more effective than providing other types of local explanations such as disclosing probabilities (see also Lundberg and Lee, 2017 and Ribeiro et al., 2016).

2.5. Linking Explainable AI and Algorithmic Trustworthiness

The main advantages of XAI mentioned in the literature are twofold: increasing user trust and preventing decision-making mistakes (Holzinger et al., 2018; Ahmad et al., 2018; Zerilli et al., 2019; Weller, 2019). The idea that XAI increases user trust is grounded in the reasoning that people will not trust systems they do not understand. Unexplained machine learning systems are highly opaque and not understandable to the decision-maker. Explaining the algorithmic functioning, which is expected to increase understandability, is therefore related to stimulating user trust (Burrel, 2016; Doran et al., 2017; Weller, 2019). Huysmans et al. (2011) empirically validated this expectation. Results of this study indicate that explanations increase the comprehensibility of algorithmic systems, which in turn was positively related to user confidence. The second function of explanations is to correct algorithmic mistakes. Ribeiro et al. (2016) demonstrated that local explanations enabled non-technical decision-makers to discern better from worse algorithmic models. Explanations consequently facilitate challenging the 'arbitrariness by algorithm' (Citron & Pasquale, 2014).

Following this reasoning, it can be hypothesized that explaining algorithmic functioning can decrease the tendency of decision-makers to rely on algorithmic advice that is incongruent with their professional judgment. This expectation is however all but certain. Van der Waa et al. (2021) demonstrate that explanations can persuade decision-makers to rely on incorrect algorithmic advice. This finding can be interpreted from a confirmation bias logic. Recall that confirmation bias describes that decision-makers tend to only search for evidence that confirms existing beliefs while disregarding other information (Jones & Sugden, 2001). Street-level bureaucrats are professional decision-makers with strong prior beliefs (Musheno & Manyard, 2003). Consequently, confirmation bias might specifically limit the effects of XAI on this type of professional frontline decision-makers.

The two functions of XAI are stimulating the perceived trustworthiness of algorithmic advice and preventing decision-making errors. To gain an understanding of the effect of XAI in complex street-level decision-making, this study assesses both these functions. I test the effect of explaining advice in general and specifically in situations where the advice is incongruent with the professional judgment of police officers. While there are some indications that XAI can increase the contestability of incongruent advice, confirmation bias logic and the recent work of van der Waa et al. (2021) render this hypothesis implausible. Rather, I hypothesize that explanations will increase the algorithmic trustworthiness of automated advice in general, but also in situations where this advice is incongruent with the professional judgment of police officers.

Two hypotheses are formulated to test for these effects. I again measure both aspects of the algorithmic trustworthiness of the advice. Hypothesis 2.1 relates explaining algorithmic advice to increased *trusting behavior* and Hypothesis 2.2 to an increase in *perceived trustworthiness*.

H2.1) Police officers are more likely to choose an option that is advised by explained algorithmic advice than an option that is advised by unexplained algorithmic advice.

H2.2) Police officers perceive explained algorithmic advice as more trustworthy than unexplained algorithmic advice.

3. EXPERIMENTAL METHODS AND MEASUREMENTS

3.1. Experimental Setting

In the introduction of this thesis, I formulated two research aims. The first aim was directed at studying the extent to which algorithmic advice induces automation bias and confirmation bias in street-level decision-making. The second aim was to research the effects of XAI on mitigating the negative effects of these two

biases. An experiment was used to test these central aims. This experiment was conducted with a population-based sample of actual street-level police officers.

Survey experiments are defined by the fact that both the experimental treatment and the outcomes are administrated in the context of a survey (Jilke & Van Ryzin, 2017, p. 118). More specifically, in this study participating police officers were, in an online survey, presented with three scenarios: a *burglary*, an *ATM robbery*, and a *stabbing incident*. Each of these scenarios involved a description of a crime that had been committed. In a dispatch report, officers were asked to help fence off the area of this crime. In this task, they were assisted by a mock algorithm. This algorithm predicted the flight routes of offenders. Participants were presented with two locations alongside predicted escape routes. The main experimental task was choosing one of these by the algorithm-assigned locations.

This experimental task was selected to appeal to a wide range of police officers. All street-level police officers have experienced – in training and mostly also in practice – this type of flight situation. Additionally, this mock algorithm was based on an algorithm that is developed by the Dutch Police organization. Both these aspects increase the ecological validity of the experimental task.

Additionally, I took great care in ensuring the ecological validity of the experimental design itself. First, I carried out multiple rounds of interviews with two police officers to understand the context and role of AI in their daily work. Specifically, I focused on understanding their reasoning and intuition when performing tasks such as fencing off the area of a crime. Furthermore, I had extensive discussions with three academic experts on the technical and socio-technical sides of AI. Based on these insights, I constructed a first experimental design.

To further strengthen this design, it was presented to a group of Ph.D. students that research the use of Artificial Intelligence within the Dutch Police, and the design was qualitatively tested with two street-level police officers. Based on the feedback of these practitioner groups, the mundane reality of the experiment was strengthened. Notably, I altered contextual information to better reflect the everyday reality of police work. Last, a small pilot study was conducted amongst a lay audience (N = 10). This indicated that the design of the experiment was clear and understandable. These extensive rounds of feedback by street-level police officers, expert groups, and the pilot study ensured that the experiment has high ecological validity.

3.2. Data Collection

Participants for this study were collected in collaboration with four regional Dutch police departments. Within each department, a contact person assisted with distributing the invitations. Not all invitations were sent on exactly the same day, but the initial invite was sent in the second week of May, and a reminder in the third or fourth week. Data collection was stopped on June 3. A total of 152 police officers responded to the survey.

Investigating the influence of algorithms on frontline work amongst a sample of actual street-level bureaucrats is important, but rare. Most research into the use of algorithms on decision-making is conducted with convenience samples (e.g. Ribeiro et al., 2016, Van der Waa et al., 2021; Alon-Barkat & Busuioc, 2021). Conducting this study with actual police officers is especially valuable considering the focus on decision-making biases and XAI. Both of which are influenced by prior knowledge of the decision-maker. Street-level bureaucrats are professional decision-makers with case-specific experience, intuition, and training (Maynard-Moody & Musheno 2003; Tummers & Bekkers, 2014). The decision-making of these professionals can be expected to differ substantially from that of lay participants. The use of a population-based sample therefore greatly enhances the generalizability and external validity of the research findings (Mullinix, Leeper, Druckman & Freese, 2015).

This focus on external validity also underlies the decision not to remove participants based on attention checks. Attention checks are a common strategy in survey experiments to increase the internal validity of results (Ejelöv & Luke, 2020). These checks have, however, been demonstrated to encourage systematic thinking (Hauser & Schwarz, 2015). This is a highly undesirable side-effect given the aim of this study to measure biases that are associated with cognitive fallacies in human thinking (Skitka et al., 1999; Kahneman, 2011). Moreover, the experimental task, fencing off the area of a crime, demands a quick response of

officers in the real world. This high cognitive load is one of the primary causes for the occurrence of automation bias (Lyell & Coiera, 2017), but cannot be replicated in a survey experiment. By including data from participants that did not fully concentrate on the experiment I measure the overall effect of the experimental conditions, i.e. the intent to treat effect. This intent to treat effect is more consistent with the effects the algorithm will have when implemented in real decision-making procedures (Hansen & Tummers, 2020).

Consequently, the only inclusion criterium was that participants completed all questions of at least one of the three scenarios. 28 participants did not satisfy this criterium. The final sample size therefore consisted of 124 police officers. The survey was distributed by local police departments, but it was estimated that the survey was sent to 400 street-level police officers. The response rate is therefore approximately 30 percent.

3.3. Experimental Design

This study presented participants with three experimental scenarios. These scenarios were described in short text fragments. Most important information was also displayed in a figure that represented a map of the area surrounding the crime (see Figure 1 for an example). The complete design of the experiment is attached as supplementary material. Here, I will specifically explain how the control group was established and subsequently highlight the two experimental treatments.

The control group in this study was used to assess the behavior of police officers when not presented with direct algorithmic advice. This group was only presented contextual information and was informed that an algorithm predicted two locations between which they were asked to choose. However, the algorithm did not recommend either of the two locations specifically. This first experimental group is therefore referred to as the *no advice* group. The police officers in this group made a choice solely on their own knowledge – these officers made a *professional judgement*.

The first experimental manipulation is the congruency of the advice with the professional judgment of police officers. In each scenario, the mock algorithm advised one location that was expected to be congruent with the professional judgment of police officers and one location that was incongruent. Experimental group 2 received *unexplained congruent* advice and experimental group 3 *unexplained incongruent* advice. This congruency of locations was determined based on the multiple rounds of qualitative interviews with police officers. For example, in the ATM-robbery scenario, as presented in Figure 1, location A is the congruent location and location B is incongruent. The rationale here was that it is common knowledge to police officers in the Netherlands that offenders of these types of crimes often flee over the highway. This could be observed by the police officers in the figure. Additionally, to further strengthen this effect, the text in this scenario informed the police officers that the offenders fled in a fast vehicle. Similarly, also the burglary and stabbing incident scenarios were designed to have one congruent and one incongruent location. Table 2 presents the most important descriptive information and congruency rationale for each scenario. While the experiment was designed to include a location that was most congruent with the professional judgment of the police officers, this was also tested post-hoc. It was measured using (1) the frequency of locations picked by the control group – the group that based their decisions solely on their professional judgment, and (2) a manipulation check.

The second experimental treatment was manipulating whether the algorithmic advice was explained. As indicated in Section 2.5, the effect of explaining advice was tested in general, and also specifically where the algorithmic advice was incongruent with the professional judgment of the police officers. Experimental group 4 consequently received *explained congruent* advice and experimental group 5 *explained incongruent* advice. The explanations provided by the algorithm were, as highlighted in Section 2.4: simple, explicit, and contrastive – explanations that provide understandability for the non-technical audience of this study. The last column in Table 2 presents a summarized translation of the explanations that were provided to the participants. The specific content of the explanations was grounded in the qualitative feedback of police officers. Whether this explanation-manipulation was successful was measured with a manipulation check.

Together the participating police officers completed 361 experimental scenarios. Figure 2 presents the complete experimental flow and division of participants over the different experimental groups.

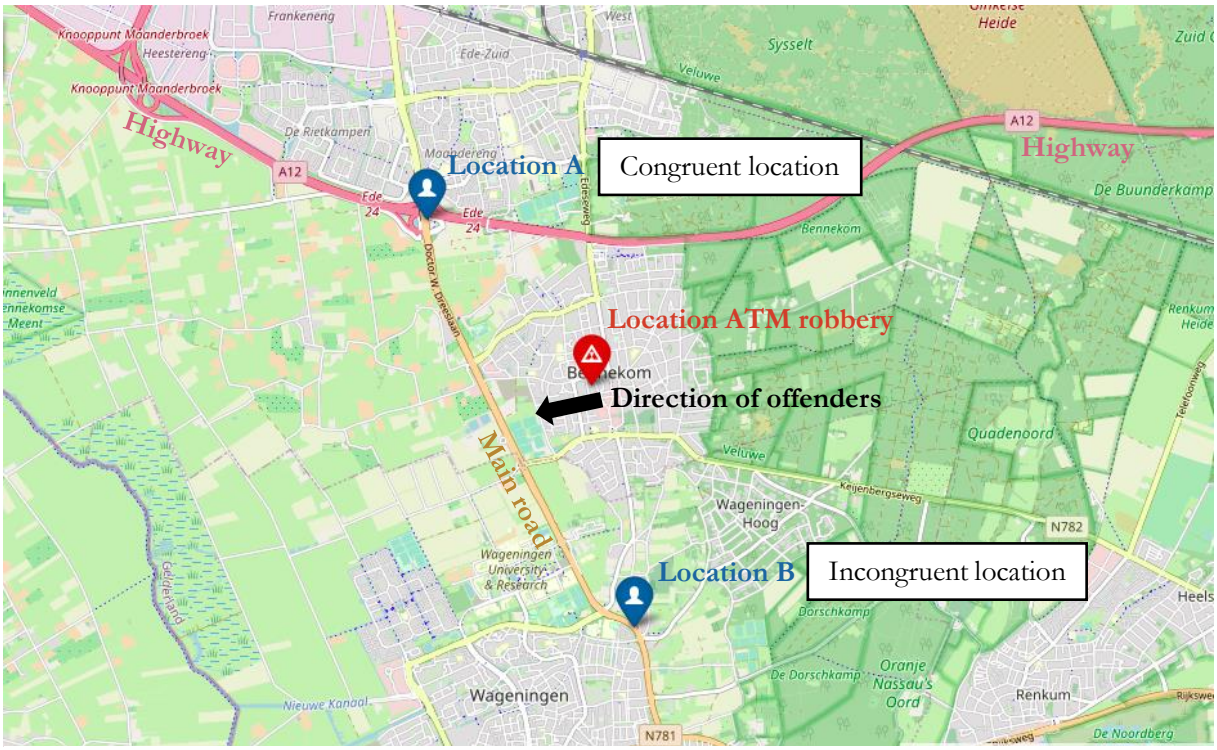


Figure 1. The map used in the ATM robbery scenario

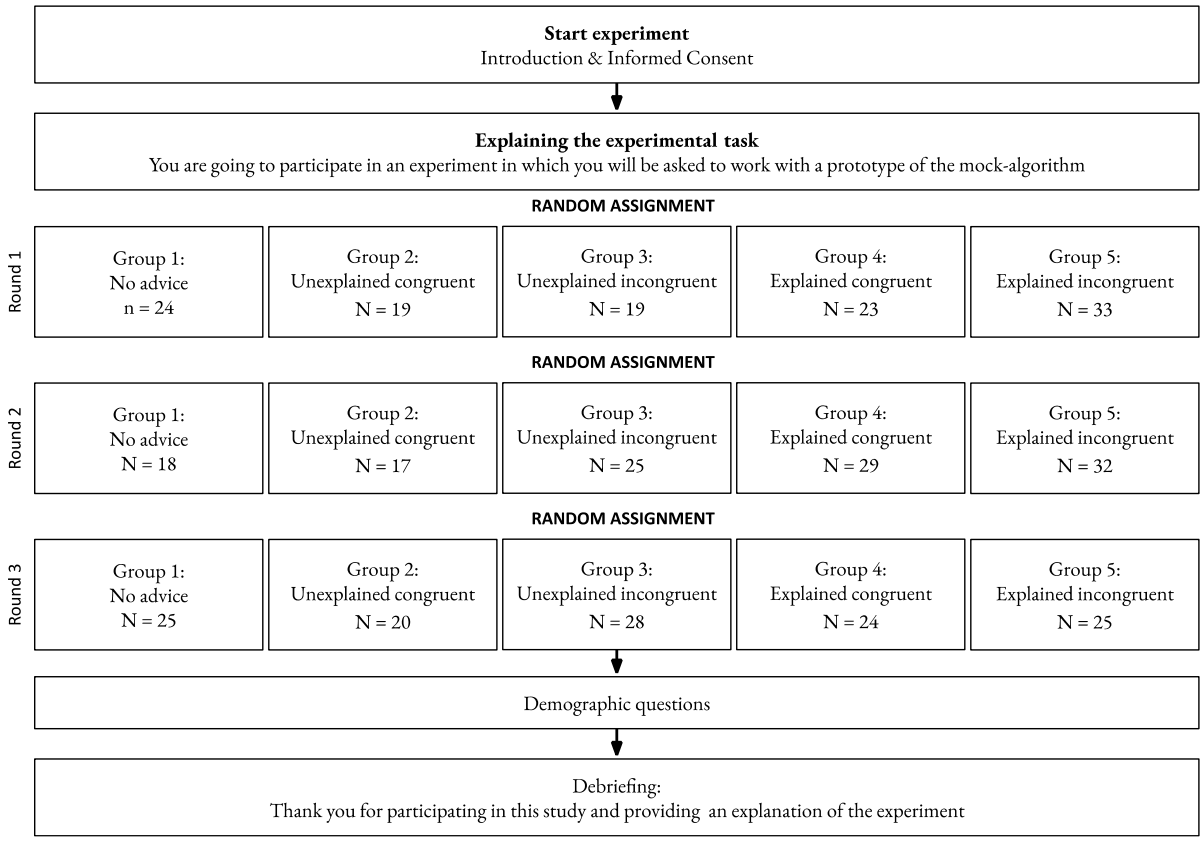


Figure 2. Flow diagram of the experiment.

Table 2. The three experimental scenarios

Scenario	Description	Congruency Rationale	Explanation Presented
Burglary	Two offenders fled north on a scooter with a German license plate.	Offenders of crimes near the border often flee abroad. The congruent location is alongside the fastest route to the German border.	Congruent location because this it is the quickest route to Germany. Incongruent location because burglary suspects often flee via quiet roads.
ATM robbery	Two offenders fled west in a fast car.	Offenders of ATM-robberies often flee via the highway. The congruent location is at the nearby highway entrance.	Congruent location because suspects of robberies often flee via the highway Incongruent location because the suspects cannot then flee into the neighboring town via this route.
Stabbing incident	The offender drove off east in a red car.	The offender has fled in a car and was last seen heading east. The congruent locations is situated east of the location where the crime has been committed crime	Congruent location because it is a central point along a major road in the direction of escape (east) from the suspect. Incongruent location because it is a central location where many possible escape routes converge.

3.4. Sample Composition

Table 3 highlights that a representative sample of street-level police officers was obtained. The average age in our sample is approximately 43, and 24% of the participants identified as females. This is comparable to the averages in the Dutch police where the average age is 45.2 and in which 34,7% of employees identify as females (Politie, 2020). More importantly, 94% of the participants indicated to have an executive status. This expresses that they are qualified to conduct street-level police work. Furthermore, most participants had post-secondary vocational education (see Appendix A). This is the education level required for street-level police work in the Netherlands. Accordingly, the obtained sample is representative of the target population.

Variables such as trust in technology and knowledge about algorithms are considered to be potentially influencing the trust of decision-makers in automated advice. In addition, demographic characteristics such as age, gender, working experience, and education level, might affect this (Alexander, Blinder & Zak, 2018). To strengthen the internal validity of the results, information regarding these variables was collected, and it was checked if the distribution of these variables over the five experimental groups was correct.

Table 3. Sample composition.

	Mean	SD
% Female	0.24	0.43
% Executive status	0.94	0.24
Years of experience*	8.78	2.42
Average Age*	43.16	10.88
Knowledge about algorithms**	2.99	1.64
Knowledge about the HOV**	2.89	1.52
General trust in technology**	5.2	1.15

*To increase anonymity age and experience were measured in ranges. The numbers presented are approximations based on group means. Age was measured on a 1 – 7 scale. Experience was measured on four different levels and therefore recoded to a dummy variable: 5 < year experience = 0, > 5 year experience = 1.

**Measured on a 1-7 scale.

The randomization checks demonstrate that there was an equal distribution of these potentially influential characteristics over the experimental groups – indicating the experimental groups are statistically comparable (see Appendix B). An ANOVA did indicate that there was a statistically significant difference in the average age between experimental groups in the stabbing incident scenario. However, since groups were not statistically significantly different in years of experience or the number of officers with an executive status, this is not expected to have a considerable impact on the research findings.

3.5. Dependent Variables

This study focused on the trust street-level decision-makers have in algorithmic advice. This was measured by investigating the algorithmic trustworthiness of the recommendation provided by the mock algorithm. Two distinct aspects of algorithmic trustworthiness were measured. First, the effect of the algorithmic recommendation on trusting behavior was investigated. This was measured by asking police officers which of the two locations they would go to intercept suspects.

The second dimension investigated was the perceived trustworthiness of the advice by police officers. This was measured using a scale developed by Grimmelikhuijsen (2019) to measure the perceived trustworthiness of algorithmic systems. This scale is based on scales that have been developed to measure trust more generally (Mayer, Davis, & Schoorman, 1995; Grimmelikhuijsen & Knies 2017), and scales designed to evaluate the trustworthiness of technological systems specifically (McKnight, Carter, Thatcher & Clay, 2011).

This perceived trustworthiness of algorithmic advice scale includes the following four Likert-scale items that all have been measured on a 1 (no trust at all) to 7 (complete trust) scale: 'I trust that the algorithm…': (1) '…used the correct information', (2) '…gave a correct decision, (3) '…assessed my situation honestly', (4) '…used all relevant information' [translated from Dutch]. To validate this scales measurements, I included a fifth item that directly asked participants whether they trusted the algorithmic prediction. These five items combined form a reliable scale to measure perceived trustworthiness (In all scenario's Cronbach Alpha above .89 and factor loadings above .6, see Appendix C).

3.6. Manipulation Checks

Two manipulation checks were used to assess whether the experimental treatments – the congruency of the advice and the effect of XAI – were successful. Additionally, the mundane reality of the scenarios was assessed. All manipulation checks were measured on a 1 (strongly disagree) to 7 (strongly agree) scale. Table 4 presents the results of these checks (descriptive statistics of these variables are presented in Appendix D).

The first manipulation check assessed the congruency of the advice with the professional judgment of police officers. This was measured by asking respondents whether they found the algorithmic recommendation they received to comply with their professional judgment. In the burglary scenario this difference was not statistically significant. As this indicates that there is no clear professional judgment, it might be more difficult to assess the occurrence of automation bias in this scenario. The other two scenarios did show a statistically significant difference, in these scenarios police officers indicated that they found the scenario that was designed to be congruent indeed more consistent with their professional judgment.

The second manipulation check assessed how participants experienced the provided explanations about the algorithmic advice. This was investigated by asking how detailed respondents perceived the advice to be. Difference tests indicate that in the burglary and stabbing incident scenarios respondents statistically significantly perceived the explained algorithmic advice to be more detailed than the unexplained algorithmic advice. No statistically significant difference was found in the ATM robbery scenario. This non-significant manipulation check cannot be explained univocally but should be kept in mind when interpreting the effects of the explanation specifically in this scenario.

Lastly, in order to obtain insight into the mundane reality of the experiment, participating police officers were asked whether they felt they could relate to the experimental scenarios. On average police officers indicated that they could somewhat agree or agree with this statement. These are good scores given that survey experiments always contain a degree of artificialness (Jilke & Van Ryzin, 2017). This experiment can

thus be considered to have a high mundane reality, which in turn strengthens the ecological validity of the research findings.

Table 4. Manipulation checks

	N	Mean	SD	Test
Burglary				
Congruence of the advice	118	4.66	1.62	t(89.30) = 1.11, p = .135, d = 0.23
Effect of explanation	118	5.05	1.26	**t(68.72) = 2.58, p = .006, d = 0.56**
Authenticity of the Scenario	118	5.12	1.31	-
ATM Robbery				
Congruence of the advice	121	4.78	1.57	**t(97.50) = 3.28, p < .001, d = 0.62**
Effect of explanation	121	4.52	1.37	t(90.47) = 1.29, p = .101, d = 0.26
Authenticity of the Scenario	121	5.30	1.17	-
Stabbing Incident				
Congruence of the advice	122	4.57	1.62	**t(86.35) = 5.78, p < .001, d = 1.13**
Effect of explanation	122	4.30	1.57	**t(94.48) = 1.73, p = .043, d = 0.35**
Authenticity of the scenario	122	5.17	1.26	-

3.7. Data Ethics and Pre-registration

In conducting this study, I followed relevant standards of data and research ethics. The data was collected via a by the Dutch Police organization supported account of the online survey tool SurveyMonkey. Privacy of respondents was warranted by only measuring the most important personal information and procedures were followed to reduce the risks of participants being identifiable, e.g. by measuring age and experience in cohorts. Prior to the experiment, participants were informed that participation was completely voluntary. Additionally, participants were briefed about the aim of the study, that this was part of academic research, and that the results would be used to write a master thesis.

Also, research ethics were taken into consideration. Most importantly, the research question, hypotheses, exclusion criteria, and measurements, were registered in the Open Science Framework (10.17605/OSF.IO/TWY9V) prior to the execution of the experiment. Small alterations from this pre-registry have to be reported. Based on a second reading of the literature, it was determined that only the general hypotheses that were pre-registered were interesting to test. Two hypotheses that were pre-registered are therefore not explicitly tested in this study.

Finally, consistent with the pre-registry, hypotheses are explored for scenarios separately but tested on the data for the three scenarios combined. This choice to combine scenarios was made to strengthen the power of the analyses and to decrease case-specific effects. The three scenarios were therefore explicitly designed to be different, but comparable. They all included the same experimental tasks – fencing off the area of a crime – and similar experimental treatments were provided. Two methodological choices were made to allow for combining the data of the different scenarios. First, to account for demand effects, scenarios were presented in random order. Second, in each scenario, respondents were randomly assigned to one of the five experimental groups. The observations are therefore considered to be independent.

4. RESULTS

This experimental study investigated the occurrence of automation bias and confirmation bias in street-level decision-making processes, and the ability of XAI to mitigate the negative effects of these biases. Both were investigated by measuring two distinct aspects that constitute the algorithmic trustworthiness of advice. In this section, I present an overview of the data. Hereafter, the effect the advice of the mock algorithm had on the behavior of police officers was tested using logistic regression analyses. Results of these analyses are presented in paragraph 4.2. Also, the perceived trustworthiness of the algorithmic recommendations was measured. This has been tested using one-way ANOVAs with planned contrast. Results of these analyses are discussed in paragraph 4.3.

4.1. Descriptive statistics of the Three Scenarios

Before proceeding with the hypothesis testing, three interesting insights into the data are highlighted. Specifically, the congruency of the advice is explored, insight is provided into the general trust police officers had in the algorithmic advice, and the effect this advice had on decision-making certainty is discussed. More elaborate descriptive statistics are found in Appendix E.

The control group in this study was used to assess whether the first experimental manipulation was successful. By assessing the choices made by this group that did not receive an algorithmic recommendation, it was measured whether the scenarios included a location that was congruent and incongruent with the professional judgment of police officers. Figure 3 presents how often the two locations were picked by this first experimental group. This highlights that in two of the three scenarios, the ATM robbery and the stabbing incident, the police officers did have a clear preference for the congruent location. In the burglary scenario, this was not the case. In this scenario, both locations were picked equally. This is consistent with the non-significant *congruency of advice* manipulation check in this scenario.

The second independent variable measured was the perceived trustworthiness of the advice. Figure 4 highlights that police officers in general have substantial trust in the algorithmic advice that was provided. Police officers on a scale from 1 (no trust at all) to 7 (complete trust), indicate that they on average found the advice to be somewhat trustworthy. While not being the direct aim of this study, the fact that police officers do not seem to be algorithmic averse in general – on the contrary even – is an interesting insight.

To get a more in-depth understanding of the decision-making process of the police officers, participating officers were asked how confident they were about the decision they made. This was measured on a 1 (not certain at all) to 100 (very certain) scale. Table 5 presents the mean certainty and standard deviations of the different groups per scenario. Notably, while the professional judgment between scenarios differs substantially, this did not influence decision-making certainty. As shown in the last row of Table 5, the type of algorithmic advice did not have a statistically significant effect on this variable.

Figure 3. Locations picked by officers in group 1

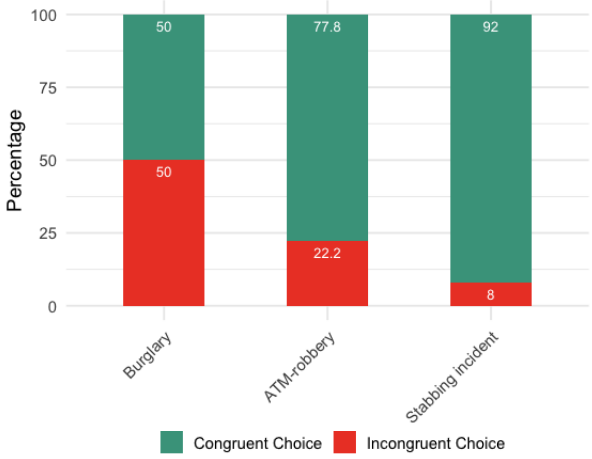


Figure 4. Mean overall perceived trustworthiness

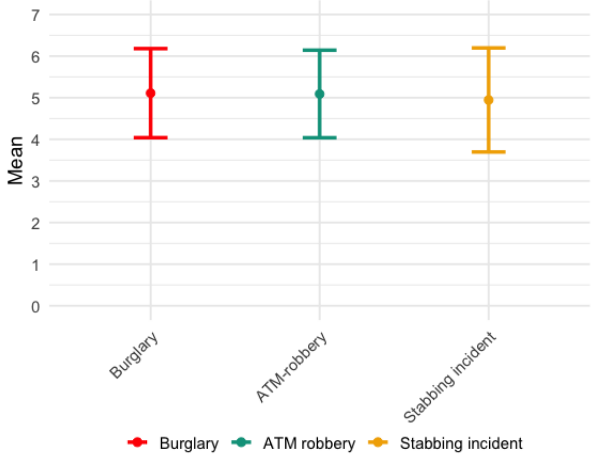


Table 5. Mean certainty of police officers in their decision-making (SD between parentheses)

	Scenario		
	Burglary	ATM-robbery	Stabbing incident
No Advice	48.1 (24.7)	53.6 (18.9)	52.0 (21.4)
Congruent	62.0 (14.8	60.5 (23.6)	54.2 (26.1)
Incongruent	47.4 (22.2)	53.1 (23.8)	49.4 (25.8)
Explained Congruent	43.9 (24.3)	60.0 (20.9)	52.3 (19.0)
Explained Incongruent	56.6 (25.9)	52.2 (29.1)	55.1 (21.1)
Difference test	F(4, 113) = 2.258, p = .0672	F(4, 116) = 0. 672, p = .613	F(4, 117) = 0.245, p = .912)

These descriptive and exploratory results provide important contextual information to interpret and explain the results of the hypothesis tested in the next two paragraphs.

4.2. The Effect of Algorithmic Advice on Behavior

Four logistic regression analyses were used to test the effects of the type of algorithmic advice on behavior. Figure 5 presents the results of these analyses (See Appendix F for the full models). This figure shows the log probability of a police officer that received algorithmic advice picking the incongruent location, compared to police officers that did not receive advice (the control group). In practice, this compares the frequency of the locations picked by police officers that only base their choice on their *professional judgment*, to the locations picked by officers that used both *professional judgment* and *algorithmic advice*. Additionally, results of these analyses can be used to discern whether explaining the algorithmic advice influenced the decision-making of police officers.

The separate analyses of the data for the three scenarios show some effects of the algorithmic advice on behavior, however, these are mostly not statistically significant. In the burglary and ATM robbery scenario no statistically significant effects were found. In the stabbing incident scenario, there is a statistically significant effect. Here, the group of police officers that received unexplained incongruent advice, statistically significantly more often picked the incongruent location than police officers in the control group. More specifically, the advice increased the probability of choosing the incongruent location by 84%. While it should be taken into consideration that the base probability of choosing this location was very low (only 8 percent of the police officers in the control group picked this location), this does indicate that in this scenario there appear to be effects of automation bias. In neither of the three scenarios separately there is evidence of confirmation bias. The congruent advice did not statistically significantly influence the decision-making of police officers. However, a general trend can be observed that congruent advice does appear to make police officers more likely to choose the congruent location.

This general trend towards confirmation bias is clearer when analyzing the data of the three scenarios combined. This analysis shows that the explained congruent advice statistically significantly altered the decision-making of police officers. It increased the probability of choosing this location by 61%. Here, incongruent advice did not have statistically significant effects on the decision-making of the participants. Consequently, police officers, in general, do not seem to be subject to automation bias. Evidence for Hypothesis 1.1 is therefore mixed; some indications are found that police officers might be subject to automation bias, but in general effects of confirmation bias appear to be strongest.

This leads to discussing the second hypothesized relation; the effects of explaining algorithmic advice on trusting behavior. Hypothesis 2.1 related providing explanations on how algorithmic advice was constructed to an increased tendency of police officers to follow this advice. Figure 5 shows that the effects of this explanation-manipulation are limited (See Appendix G for the full model). While especially in the ATM robbery and stabbing incident, XAI seems to have some effect on increasing the contestability of incongruent advice, these effects are small and are not statistically significant. While this study does not completely render out the effect of XAI on trusting behavior, these effects are at best limited and not detectable by the relatively small sample size of this study. This absence of effect of providing explanations was confirmed by the analysis of the combined data. Here, hardly any variation between groups that received unexplained and explained advice is visible. Based on these results Hypothesis 2.1 thus can be rejected. No evidence is found that police officers are more likely to follow explained algorithmic advice than unexplained algorithmic advice.

The previous analyses revealed that the explanation of advice had, at best, only a very minor influence on the behavior of police officers. More exploratory, the responses of police officers in the unexplained and explained advice groups – given that the effect of this manipulation on behavior was insignificant – were combined. This results in three experimental groups: a *no advice* group, a *congruent advice* group, and an *incongruent advice* group. For each scenario and the scenarios combined, a logistic regression analysis was again used to compare the behavior of the two groups that received algorithmic advice to the behavior of the no advice group. Figure 6, demonstrates the results of these analyses. This again shows that the effect of the congruent advice seems to be larger than the effects of the incongruent advice. Most notably, the logistic regression for data of the scenarios combined shows that congruent algorithmic advice had a

statistically significant influence on decision-making. It increased the probability of picking this incongruent location by 59%. No statistically significant effect of the incongruent advice was found. These results confirm the previous conclusions regarding Hypothesis 1.1. It highlights that confirmation bias is likely to have a stronger effect on how police officers validate algorithmic outputs than automation bias.

4.3. The Perceived Trustworthiness of Algorithmic Advice

The second aspect of algorithmic trustworthiness that was assessed in this study was the trusting attitude of police officers. This was measured using a scale that combined five questions which together reliably measured the perceived trustworthiness of the advice provided by the mock algorithm. Figure 7 presents the means and confidence interval of this combined perceived trustworthiness scale of each experimental group for the three scenarios separately and for the scenarios combined.

Hypothesis 1.2 expected that perceived trustworthiness depended on whether the algorithmic advice was congruent or incongruent with the general professional judgment of police officers. In each of the scenarios, a one-way ANOVA with planned contrast tested this effect (see Appendix H for a more detailed explanation of this estimation strategy). In none of the analyses on the data of the three scenario separately a statistically significant effect was found (Burglary: $F(4, 113) = 2.11$, $p = .0842$, ATM-robbery: $F(4, 116) = 0.81$, $p = .524$; Stabbing incident: $F(4, 117) = 1.62$, $p = .174$).

Figure 5. The mean and 95% confidence interval of the logistic regression estimators

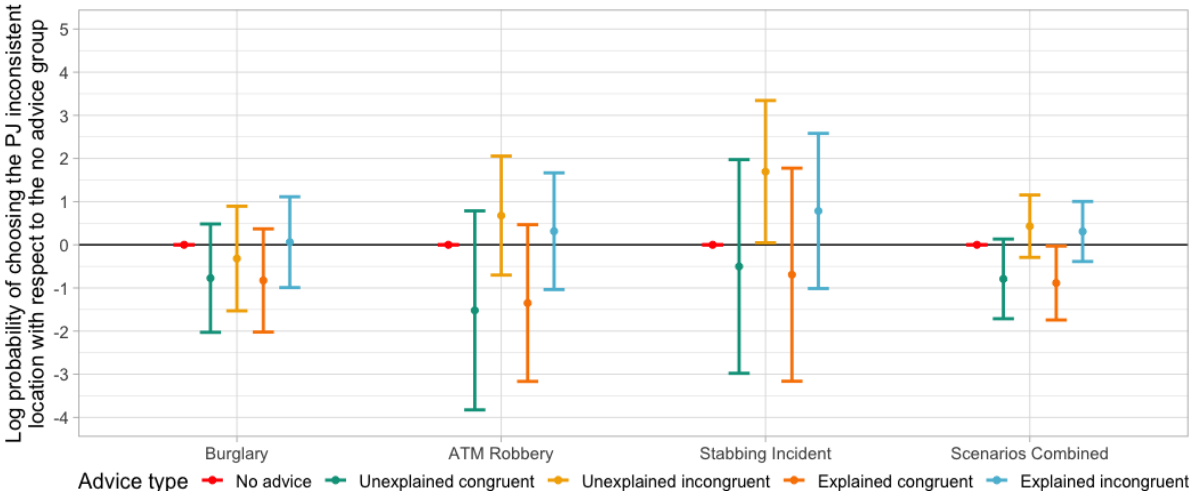
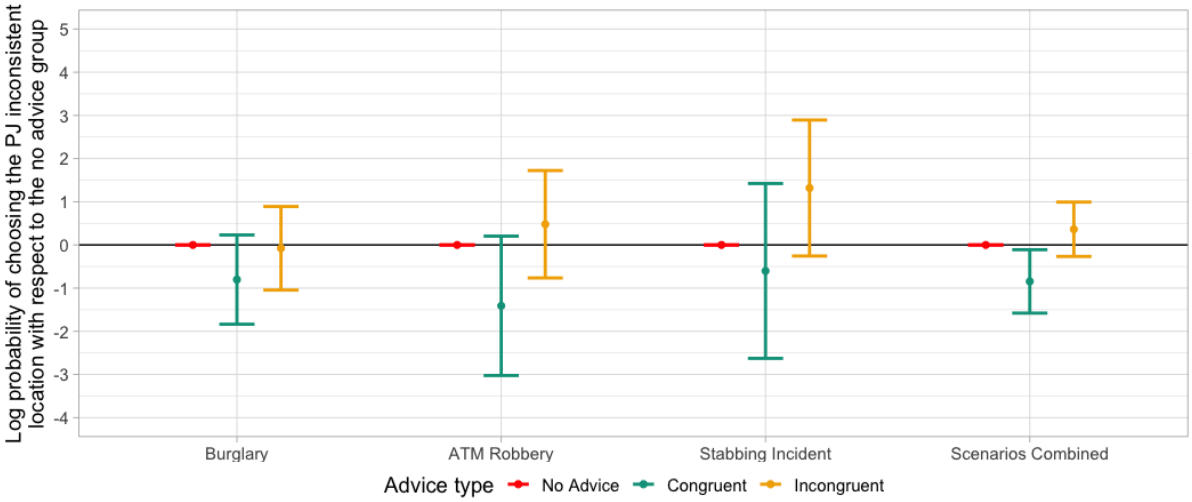
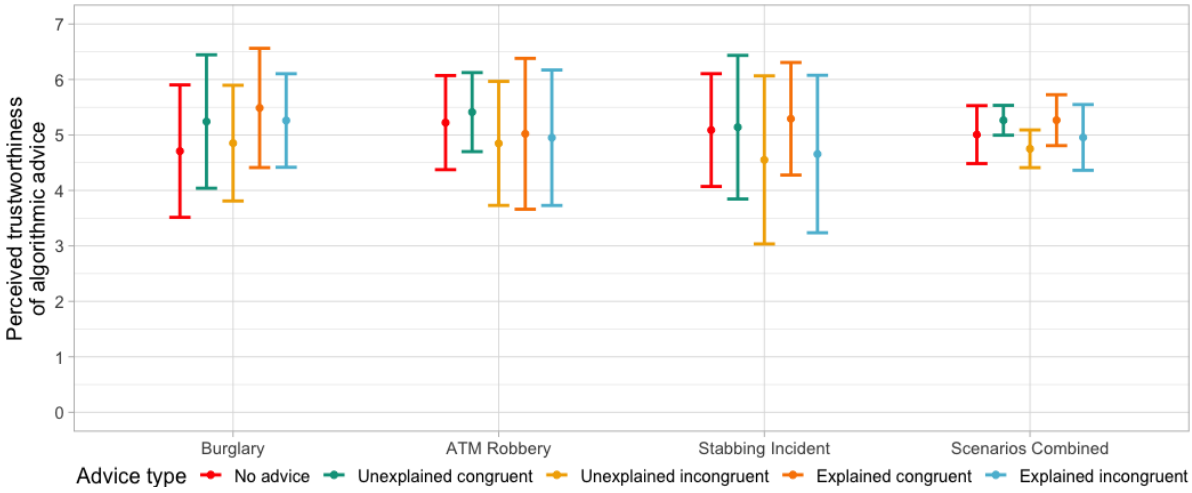


Figure 6. The mean and 95% confidence interval of the logistic regression estimations for the explained and unexplained advice combined.



Note: In both plots, confidence interval not intersection with the zero-line indicates a statistically significant difference of the log-probability of the groups that received advice with respect to the no advice group (red)

Figure 7. The mean and 95% confidence interval of perceived trustworthiness scale



A similar analysis tested for these effects in the combined data. This ANOVA was statistically significant ($F(4, 356) = 2.44$, $p = .047$). The full model specifications can be found in Appendix I. Most prominently, a planned contrast demonstrates that police officers perceived algorithmic advice that is congruent with their professional judgment to be more trustworthy than advice that was incongruent. This effect is statistically significant and can be considered to be quite large ($t(4, 356) = 2.53$, $p = .012$, $\eta 2 = 0.13$). Hypothesis 1.2 can therefore be accepted. This finding is in accordance with the conclusion on the behavioral dimension of algorithmic trustworthiness; it also indicates that police officers are subject to confirmation bias rather than to automation bias when interpreting algorithmic output.

The planned contrasts in the ANOVA combined furthermore lead to rejecting hypothesis 2.2. The results highlight that there is no difference between the perceived trustworthiness of explained and unexplained algorithmic advice. This effects was found in general ($t(4, 356) = -0.06$, $p = 0.192$), but also specifically for advice that was incongruent with the prior beliefs of the police officers ($t(4, 356) = -0.129$, $p = .012$). This is consistent with the findings on trusting behavior, which was also not or only to a limited extent influenced by whether the algorithmic advice was explained or not.

4.4. Algorithmic Biases in Decision Making and the Effects of Explainable AI

The results of this study are open to interpretation. Using the descriptive information provided in the first paragraph of this section and through combining insights from the different analyses some interesting patterns do emerge.

First, the results presented imply that police officers selectively confirmed the algorithmic advice. Congruent advice did statistically significantly influence the decision-making of police officers, while incongruent advice, in general, did not. Evidence for hypothesis 1.1 is therefore mixed. The second aspect of algorithmic trustworthiness – the perceived trustworthiness of advice – helps to explain this finding. With the acceptance of Hypothesis 1.2, I demonstrated that police officers perceived the congruent advice to be more trustworthy than the incongruent advice. These factors combined are a convincing indication that police officers are subject to confirmation bias rather than to automation bias when interpreting algorithmic outcomes.

Second, the results imply that explaining how the algorithmic advice was constructed did not influence the algorithmic trustworthiness of the advice. While some effects of the explanations were visible, these were rather limited. The logistic regression analysis showed that there was no effect of explaining the advice on trusting behavior and the ANOVA's demonstrated no statistically significant effects of manipulating this variable on the perceived trustworthiness of the advice either. Both hypotheses regarding the effects of XAI, Hypothesis 2.1 and 2.2, were rejected.

Table 6 summarizes the main findings. These increase our understanding of how algorithms influence street-level decision-making. More specifically, in the next section, the contributions to public administration and XAI literature are highlighted. Additionally, I discuss the practical implications of the research findings.

Table 6. Summary of the findings

Concept	Hypothesis	Finding
Automation Bias & Confirmation Bias	H1.1) Police officers are more likely to choose an option when this is advised by an algorithm compared to making the same choice based solely on their professional judgement.	Mixed Evidence
	H1.2) Police officers perceive algorithmic advice that is congruent with their professional judgment as more trustworthy than algorithmic advice that is incongruent with their professional judgment.	Accepted
Explainable AI (XAI)	H2.1) Police officers are more likely to choose an option that is advised by explained algorithmic advice than an option that is advised by unexplained algorithmic advice.	Rejected
	H2.2) Police officers perceive explained algorithmic advice as more trustworthy than unexplained algorithmic advice.	Rejected

5. DISCUSSION

Machine learning algorithms increasingly infer in complex street-level decision-making processes (Young et al., 2019; Bullock, 2019). This raises concerns, the use of algorithms is linked to inducing new decision-making biases (Mosier & Skitka; 1996; Alon-Barkat, 2021). This research explicitly focused on two of these possible biases induced by algorithms: automation bias and confirmation bias. Providing explanations about how algorithmic advice is constructed, using explainable AI (XAI), is often described to mitigate negative effects associated with these biases (Ribeiro et al., 2016; Ahmad et al., 2018; Weller, 2019). These expectations formed the basis for the two aims central in this study: investigating algorithm-induced decision-making biases and the effects of XAI on mitigating the negative effects of these biases. Specifically, an answer is formulated to the research question: *Does algorithmic advice introduce automation bias and confirmation bias in street-level decision-making, and can explainable AI help to mitigate the negative effects of these biases?*

This question is answered using a survey experiment conducted amongst a sample of frontline police officers. In the experiment, a mock-algorithm assisted officers in the street-level task of fencing off the area of a crime. In three different but comparable scenarios: a burglary, an ATM robbery, and a stabbing incident, the trust of police officers in algorithmic advice provided by this mock-algorithm was assessed. The results of this experiment imply that (1) street-level bureaucrats are *not* prone to automation bias, rather (2) they *are* likely to be subject to confirmation bias. Additionally, this study finds that (3) the effects of explaining algorithmic advice might only be *limited* for professional decision-makers. These findings have important implications for how street-level decision-making processes can be enabled by algorithms.

The first core message of this research is that risks posed by automation bias in complex street-level decision-making seem to be small. Two of the three scenarios and the analysis of the data combined demonstrated police officers deferred algorithmic advice that was incongruent with their professional judgment. The risks of automation bias in complex street-level decision-making processes thus appear to be much less prominent than found in decision-making tasks in for example aviation and public health (Skitka et al., 1999; Lyell & Coiera, 2017).

The second core message is that the results imply that street-level bureaucrats are subject to confirmation bias. This was demonstrated by the finding that, in general, the congruent algorithmic advice did affect police officers' decision-making. Furthermore, this indication of the occurrence of confirmation bias was strengthened by the acceptance of Hypothesis 1.2, which showed that police officers perceive incongruent algorithmic advice to be less trustworthy than congruent advice. In accordance with Alon-Barkat and

Busuioc (2021), I therefore also conclude that confirmation bias seems to have a stronger effect on how street-level decision-makers interpreted algorithmic outputs than automation bias.

The third core message is that in complex public decision-making processes the ability of XAI to prevent algorithm-induced decision-making errors appears to be limited. Prior research in the field of XAI is empirically limited and inconclusive (Adadi & Berrada, 2018). Van der Waa et al. (2021) for example demonstrated that explanations increased the tendency of decision-makers to follow algorithmic advice, even when this advice is wrong. Other studies showed that explanations increase the ability of decision-makers to spot algorithmic mistakes (Ribeiro et al., 2016). This study found neither of these effects. The results indicate that explaining how algorithmic advice is constructed does not influence trusting behavior or the perceived trustworthiness of this advice.

These three core messages hold important contributions to theory. More specifically, in the following two paragraphs, I discuss implications for the literature on the use of algorithms in the street-level bureaucracy and the significance of the findings for XAI literature. Subsequently, paragraph 1.3 elaborates on the practical implications of the research findings and paragraph 1.4 discusses two methodological limitations of this study. Paragraph 1.4 presents a general conclusion and future outlook.

5.1. Implications for Street-level Decision-Making

The role of ICT in street-level decision-making processes has been viewed from both an enablement and a curtailment thesis (Buffat, 2015; Busch & Henriksen, 2018). The enabling perspective describes that algorithms help overcome weaknesses in administrative discretion. Algorithms can however also curtail if they overly restrict the exercise of administrative discretion. The results of this study mostly support the enabling view.

This enabling function is grounded in the finding that street-level bureaucrats do not seem to be subject to automation bias, but are prone to confirmation bias. Algorithms can make mistakes and frontline work demands case-by-case judgment (Binns, 2020). The street-level bureaucrat is therefore kept in-the-loop of a decision-support system to provide individual judgment and prevent algorithmic mistakes (Veale & Brass, 2018; Peeters & Widlak, 2018; van Eijk, 2020). Automation bias results in algorithmic advice always prevailing over the professional judgment of the street-level bureaucrat, even when faced with contradictory evidence (c.f. Skitka et al., 1999). This means that, if street-level bureaucrats are subject to automation bias, individual judgment is not be provided and algorithmic errors are not prevented. In sum, automation bias renders keeping a human-in-the-loop a 'moot point' (Peeters, 2020).

The fact that no strong evidence for automation bias was found indicates that street-level bureaucrats can control algorithmic advice. This is likely the effect of police officers being subject to confirmation bias. According to confirmation bias theory, decision-makers only use additional information if this congruent with their prior beliefs (Nickerson, 1998; Klayman, 1995). This research showed that police officers are prone to this type of selective adherence when interpreting algorithmic outcomes. Police officers do incorporate algorithmic advice in their decision-making process, however only if this advice was to some extent congruent with their prior beliefs. Confirmation bias from this perspective thus leads to a selective adoption of algorithmic advice which preserves the value of professional judgment and administrative discretion.

This finding – confirmation bias rather than automation bias – can be understood by looking at the nature of frontline work. In their daily work, street-level bureaucrats are accustomed to translating general rules to fit the local situations they encounter (Lipsky, 1980; Tummers & Bekkers, 2014). This translation is guided by their views of appropriate action (Maynard-Moody & Musheno 2003; Bannink; 2018). The results of this study imply police officers interpret algorithmic advice as a new type of general regulation, instead of seeing it as a direct decision-making aid. This study indicates that street-level bureaucrats will use algorithmic information, but as with any general regulation, only selectively. They do not trust the algorithmic advice if it is not adapted to the local situation. This function of algorithmic advice can, following de Boer and Raaphorst (2021), be seen as an *algorithmic nudge*. By analyzing large amounts of data algorithms provide decision-makers with generalizable and consistent information (Busch & Henriksen,

2018). Algorithmic nudges can direct the decision-makers towards this desired generalizable decision-making if required, but also leave room for the local adaptation of this general advice.

While this presents a mostly optimistic view on the effects of confirmation bias, the occurrence of this bias also has negative implications. Confirmation bias also implies that street-level bureaucrats will not be able to correct algorithmic advice that is consistent with their prior beliefs. This poses new risks. Machine learning algorithms are prone to reproduce unfair biases in human decision-making (O'Neil, 2016; Diakopoulos, 2016). As shown by Alon-Barkat and Busuioc (2021), confirmation bias can then also lead to algorithms nudging decision-makers towards prejudicial decision-making. This undermines the effectiveness of regulations that require a human-in-the-loop to prevent unfair algorithmic outcomes (e.g. Wagner; 2016; Jung et al., 2019; European Commission, 2021).

Additional research is therefore required to investigate how algorithmic nudges can best be applied. Three aspects are especially important. First, while automation bias might not be present in street-level decision-making at this moment, this does not mean it never will. Research on automation bias has mostly been performed in fields that are automated by highly reliable systems such as aviation (Mosier et al., 1998; Lyell & Coiera, 2017). It is much harder for decision-makers to control and contest algorithms that have been proven to be almost always reliable (Peeters, 2020). Future research should therefore investigate how repeated use of reliable algorithmic systems affects the occurrence of decision-making biases. Secondly, one of the major strengths of this study is that it was conducted within a population-based sample of street-level police officers. Different types of street level-bureaucrats are to some extent comparable. They are all frontline workers that directly interact with citizens and that are accustomed to having substantial discretion (Lipsky, 1980). The findings of this study can therefore probably be generalized to other types of street-level bureaucrats – at least to a certain degree. However, groups of street-level bureaucrats also have distinct characteristics (Maynard-Moody & Musheno, 2003). Conducting similar experimental studies with other types of street-level bureaucrats, such as nurses and teachers, is required to assess how different frontline workers use algorithmic advice in their work. Thirdly, it should be studied how the design of algorithmic systems affects the occurrence of decision-making biases. For example, the mock algorithm used in this experimental study presented police officers with advice but left room for administrative discretion by presenting two options. It is important to explore how this type of design choices affect the use of algorithmic systems.

In this research one design elements that in literature is often indicated to enable the correct use of algorithm advice – explaining the functioning of the algorithm – was tested. The next paragraph discusses how the findings of this study contribute to literature on XAI.

5.2. Implications for Explainable AI

The second contribution is to the literature on XAI. This study tested for the effects of everyday, contrastive, and simple explanations. This type of explanation was expected to be highly effective (Miller, 2019). Findings in this study are nonetheless sobering. They indicate that the provided explanations had no or only minor effects on the trust of police officers in the algorithmic advice. In doing so, this study provides a counterbalance to the high expectations of explaining how algorithms function in the literature (e.g. Ribeiro et al., 2016; Holzinger et al., 2018; Zerilli et al., 2019; Weller, 2019).

This conclusion, that explanations do not seem to affect how street-level bureaucrats use algorithmic advice, is consistent with the earlier finding in this thesis that frontline workers are subject to confirmation bias. Professional decision-makers, such as street-level bureaucrats, have a high degree of prior knowledge. Being subject to confirmation bias, the professional-decision maker is therefore unlikely to change its view based on explanations that contrast their own reasoning. This, at least to some extent, also explains the difference between the results of this study and findings in prior research (e.g. van der Waa et al. 2021, Ribeiro et al. 2016). These authors all studied the effects of XAI within a sample of lay participants. Lay-audiences vis-à-vis street-level bureaucrats have different prior knowledge and beliefs. This shows the value of experimentally testing assumptions from generic literature in applied and complex contexts.

The value of XAI in complex decision-making processes should, however, not be disregarded based solely on the results of this study. The results did in fact indicate that explaining algorithmic functioning might

have some effects – especially on increasing contestability of the incongruent advice. These observed effects were however small and statistically not significant. Furthermore, only a specific type of explanation was tested. This can be another cause of the differences between the findings in this study and earlier work on the effects of XAI. More research into XAI is therefore necessary. Especially interesting is testing one of the most important aspects of XAI systems: the ability to increase contestability of algorithmic advice when the algorithm errs (Ribeiro et al., 2016; Ahmad et al., 2018; Adadi & Berrada, 2018). While this study shows that more generic explanations might have little effect, future XAI studies should aim to research the effects of explanations that explicitly or inexplicitly demonstrate algorithmic mistakes, and how this is affected by confirmation bias.

Notably, explaining how algorithms function is necessary for more than only preventing decision-making mistakes. Explanations are required to strengthen citizen's trust and to ensure public organizations that use machine learning approaches can be held accountable (Meijer & Grimmelikhuijsen, 2020). Research into XAI techniques therefore remains necessary, even while the effects of these systems on improving decision-making quality might be limited.

5.3. Implications for Practice

This thesis is one of the first studies that empirically researched the effects of algorithms on street-level decision-making. The results of this research, therefore, do not lend themselves to make direct practical recommendations. The research does provide a general insight into the required condition under which algorithms can enable frontline work. Additional experimentation with algorithms in the street-level bureaucracy is however required to go from this general insight towards practical guidance. This thesis can serve as a trigger for these future studies.

The foremost practical contribution of this thesis is providing empirical evidence for the enabling function of algorithms within street-level work (Busuioc, 2020; Young et al., 2019). This research showed that algorithmic nudges potentially can steer decision-makers towards desired behavior. However, as previously highlighted by Alon-Barkat and Busuioc (2021), algorithmic nudges can also direct street-level bureaucrats towards undesired behavior. In general, it can therefore be stated that algorithmic advice needs to be strong enough to nudge decision-makers towards desired behavior, but simultaneously leave room for administrative discretion. Additional research is warranted to find this balance between algorithmic and administrative discretion.

The second practical contribution of this thesis lies in showing that this additional research is desirable and possible. Algorithms help overcome weaknesses in human decision-making, e.g. by increasing decision-making effectiveness, consistency, and generalizability (Le Sueur 2015; Brundage et al., 2018; Bullock, 2019; Young et al., 2019; Binns, 2020). Machine learning approaches are, however, also linked to increasing the unfairness of public decision-making processes (Peeters, 2020; van Eijk, 2020). Public organizations can therefore be hesitant to experiment with the use of algorithms. The findings in this research indicate that, even if an algorithm provides incorrect advice, street-level bureaucrats will in many cases be able to control and correct this. Experimentations with 'less-than-perfect' automation do therefore not automatically lead to decision-making errors in the street-level bureaucracy (c.f. Lyell & Coiera 2017). This is valuable information. It shows that testing the effects of algorithms in practice is possible. Experimentation with algorithms in survey experiments, as well as in practice, are the only approach to consolidate the promises that algorithms bring to frontline work.

5.4. Methodical Limitations

As with any study, this research is subject to methodological limitations. The first is the limited sample size. Post-hoc power analysis demonstrates that the power obtained in the stabbing incident scenario – the scenario in which effects were most prominent – was 0.68. This indicates that the null effects in other scenarios, both for the effects of automation and confirmation bias, could be the result of limited power. Similarly, while the results indicate some minor effects of providing explanations, these small effects were not statistically significant. In general, post-hoc power analyses indicate that only if the explanation manipulation changed the decision-making of at least 15% of police officers, this study had a reasonable chance of measuring this difference (to obtain a power of at least 0.80). Being one of the first studies investigating these processes within a sample of real-decision makers, these are valuable insights. However,

the scalability of algorithms means that they can be implemented in many decision-making procedures at the same time. Small differences, as a result, can have big consequences. Future research should aim to research by automation-induced biases and effects of XAI techniques with larger groups of decision-makers.

The second methodological limitation is the aggregation of the data of the three scenarios. To account for the limited power, this research combined data of the three scenarios and analyzed these as if the observations were independent. This procedure could be justified because the same treatments were provided in comparable scenarios. Additionally, the scenarios were presented in random order, and participants were in each scenario randomly distributed over the experimental groups. Furthermore, participants were specifically indicated to not let scenarios influence each other. The analyses of the aggregated data confirmed the trends in the separate scenarios. Combined this study therefore provides valid evidence that the effect of confirmation bias seems to be stronger than automation bias and that XAI is likely to only have a limited effect. Nonetheless, the influence of demand and subject level effects, and the effects of small size differences per experimental group, cannot be ruled out completely. This possible dependency within the data has to be kept in mind when interpreting the results of the analyses on the data of the combined scenarios. Future experimental studies should aim to validate these results with truly independent observations.

5.5. Conclusion

With the increase of data available to organizations, machine learning approaches are increasingly able to infer in complex public decision-making tasks (Meijer, 2018, Young et al., 2019). Algorithmic systems might enable or curtail frontline work (Buffat, 2015; Bush & Henriksen, 2018). However, insights into the impact that machine learning algorithms have on street-level decision-making processes are limited in terms of empirical evidence (Bullock et al., 2020; Gritsenko & Wood, 2020). An increased understanding of how algorithms influence public decision-making is therefore needed. This study focused on an essential element in the use of algorithms by public organizations: the expertise of the street-level bureaucrat (Meijer & Grimmelikhuijsen, 2020). Street-level bureaucrats are the public sector workers that directly interact with citizens (Lipsky, 1980). The effect algorithms have on the decision-making of these public sector workers consequently has a direct impact on how public services are delivered.

This study shows that it might be the very nature of street work itself that provides street-level bureaucrats with the expertise essential to work with predictive algorithmic systems. Street-level bureaucrats on a regular basis translate general rules to apply in local situations (Lipsky, 1980; Maynard-Moody & Musheno, 2003; Tummers & Bekkers, 2014). This study demonstrated that this adaptation of general rules to specific circumstances can also enable the correct use of algorithmic information. The findings highlight that street-level bureaucrats are likely to selectively adopt algorithmic advice, especially when this is congruent with their professional judgment. It is precisely this constructive but also critical attitude that is required for the successful implementation of algorithmic systems. This emphasizes that under the right conditions – where administrative and algorithmic discretion are both present – algorithms can enable frontline work.

Understanding algorithmic functioning, obtained through XAI, was expected to be a fundamental enabling condition (Rader et al., 2016; Ananny & Crawford, 2018; Bannister & Connolly, 2020). This research demonstrated this expectation has to be revisited. While explanations are needed for more than the controllability of algorithmic advice, the absence of effect of XAI emphasizes that the adoption of algorithmic systems requires more than a technical perspective (Meijer & Grimmelikhuijsen, 2020).

This research contributed to this wider perspective by focusing on how street-level bureaucrats use algorithmic advice. The impact of algorithms on frontline work is, however, affected by organizational choices that transcend the decision-making of the frontline worker. New policies are developed to guide how algorithmic advice should be used, training programs inform decision-makers how to work with algorithmic systems, and new collaborative structures between system designers and users take shape. Moreover, once the algorithm is implemented, new monitoring systems are put in place to evaluate algorithmic outcomes. These factors ultimately all affect how public services are delivered (Meijer & Grimmelikhuijsen, 2020). A holistic view of the use of algorithms in public organizations is therefore essential. This view should look at machine learning itself, but also encompass observing how algorithms

alter social, managerial, and organizational processes (Alexander et al., 2018; Meijer et al., 2021). This study then also reiterates the need for more insights into the influence of algorithms on public decision-making from a public administration perspective. These additional insights are essential to realizing the promises that algorithms have to offer for street-level work.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138-52160.

Ahmad, M. A., Teredesai, A., & Eckert, C. (2018). Interpretable machine learning in healthcare. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 447–447.

Alexander, V., Blinder, C., & Zak, P. J. (2018). Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior*, *89*, 279-288.

Alon-Barkat, S., & Busuioc, M. (2021). Decision-makers Processing of AI Algorithmic Advice: Automation Bias versus Selective Adherence. *arXiv preprint arXiv:2103.02381*.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989.

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135). Pergamon.

Bannink, D. B. D. (2018). Implementation management: The work and social assistance act. In H. Boutellier, & W. Trommel (Eds.), *Emerging governance: Crafting communities in an improvising society* (pp. 119-134). Eleven International Publishing.

Bannister, F., & Connolly, R. (2020). Administration by algorithm: A risk management framework. *Information Polity*, *25*(4), 471-490.

Battaglio, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral public administration *ad fontes:* a synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. *Public Administration Review*, *79*(3), 304–320.

Bayamlıoğlu, E. (2021). The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation". *Regulation & Governance*.

Binns, R. (2020). Human Judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, rego.12358.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.

Buffat, A. (2015). Street-level bureaucracy and e-government. *Public Management Review*, *17*(1), 149–161.

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

Bullock, J., Young, M. M., & Wang, Y.-F. (2020). Artificial intelligence, bureaucratic form, and discretion in public service. *Information Polity*, 1–16.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251.

Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity*, *23*(1), 3–28. https://doi.org/10.3233/IP-170050

Busuioc, M. (2020). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, puar.13293.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Wash. L. Rev.*, *89*, 1.

Davis, K. C. (1969) Discretionary Justice: A Preliminary Inquiry, Baton Rouge, LA: Louisiana State University Press.

de Boer, N., & Raaphorst, N. (2021). Automation and discretion: Explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review*, 1–21.

Dechesne, F., Dignum, V., Zardiashvili, L., & Bieger, L. J. (2019). AI & Ethics at the Police.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62.

Doran, D., Schulz, S., & Besold, T. R. (2017). What does explainable AI really mean? A new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794*.

Ejelöv, E., & Luke, T. J. (2020). "Rarely safe to assume": Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937.

European Commission. (2021). Proposal for a Regulation of the European parliament and of the council: Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. Brussels.

Giest, S., & Grimmelikhuijsen, S. (2020). Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective. *Information Polity*, *25*(4), 409–417.

Grimmelikhuijsen, S. (2019). *Deciding by algorithm: testing the effects of algorithmic (non-) transparency on citizen trust* [Paper presentation]. EGPA 2019 Belfast, Northern Ireland.

Grimmelikhuijsen, S., & Knies, E. (2017). Validating a scale for citizen trust in government organizations. *International Review of Administrative Sciences*, 83(3), 583-601.

Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. *Public administration review*, *73*(4), 575-586.

Gritsenko, D., & Wood, M. (2020). Algorithmic governance: A modes of governance approach. *Regulation & Governance*, rego.12367.

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

Hansen, J. A., & Tummers, L. (2020). A systematic review of field experiments in public administration. *Public Administration Review*, *80*(6), 921-931.

Hauser, D. J., & Schwarz, N. (2015). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *Sage Open, 5*(2), 2158244015584617.

Holzinger, K., Mak, K., Kieseberg, P., & Holzinger, A. (2018). Can we trust machine learning results? artificial intelligence in safety-critical decision support. *ERCIM NEWS*, (112), 42-43.

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., & Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*, *51*(1), 141-154.

Jilke, S. R., & Van Ryzin, G. G. (2017). Survey experiments for public management research. In O. James, S. R. Jilke, & G. G. Van Ryzin (Red.), *Experiments in Public Management Research* (pp. 117–138). Cambridge University Press.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, *50*(1), 59-99.

Jung, C., Mueller, H., Pedemonte, S., Plances, S., & Thew, O. (2019). Machine learning in UK financial services. *Bank of England and Financial Conduct Authority*.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of applied research in memory and cognition*, *2*(1), 42-52.

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

Kizilcec, R. F. (2016, May). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390-2395).

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, *32*, 385-418.

Kramer, R. M., & Lewicki, R. J. (2010). Repairing and enhancing trust: Approaches to reducing organizational trust deficits. *Academy of Management annals*, *4*(1), 245-277.

Lipsky, M. (1980). Street Level Bureaucracy: Dilemmas of the Individual in Public Services. Russell Sage Foundation.

Lipton, P. (1990). Contrastive explanation. *Royal Institute of Philosophy Supplements*, *27*, 247-266.

Lundberg, S., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. Journal of the American Medical Informatics Association, 24(2), 423–431.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709-734.

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

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)*, *2*(2), 1-25.

Meijer, A. (2018). Datapolis: a public governance perspective on "smart cities". *Perspectives on Public Management and Governance*, *1*(3), 195-206.

Meijer, A., & Grimmelikhuijsen, S. (2020). How to generate citizen trust in governmental usage of algorithms. In M. Schuilenburg & R. Peeters (Eds.), *The algorithmic society: Technology, power, and knowledge* (pp. 53-66). Routledge.

Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, *42*(12), 1031–1039.

Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. Public Administration Review, puar.13391. https://doi.org/10.1111/puar.13391

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors*, *58*(3), 401-415.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In: R Parasuraman, M Mouloua, (Eds.) Automation and Human Performance: Theory and Applications. Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1996:201–220.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology*, *8*(1), 47–63.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*(2), 109-138.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175-220.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. London: Penguin.

Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, 1–16.

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

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. Nature Human Behaviour, 4(7), 736–745.

Rader, E., Cotter, K., & Cho, J. (2018, April). Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-13).

Rathenau Instituut. (2019, June 13). Dankzij deze sensoren kunnen rondreizende bandieten minder hun gang gaan. Rathenau Instiuut. Retrievable from:

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*(3), 429.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, *23*(3), 393-404.

Simon, H. A. (1957). *Models of man; social and rational.* Wiley.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006.

Thagard, P. (1989). Explanatory coherence. Behavioral and brain sciences, 12(3), 435-502.

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., & Marsch, S. U. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research*, *40*(3), 271-300.

Tummers, L., & Bekkers, V. (2014). Policy implementation, street-level bureaucracy, and the importance of discretion. Public Management Review, 16(4), 527–547.

van der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence*, *291*, 103404.

Van Eijk, G. (2020). Algorithmic reasoning: The production of subjectivity through data. In: Schuilenburg, M., & Peeters, R. (eds.), *The Algorithmic Society: Power, Knowledge and Technology in the Age of Algorithms*. London: Routledge.

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In: Karen Yeung and Martin Lodge (eds.): *Algorithmic Regulation*. Oxford: Oxford University Press.

Wagner, B. (2016). Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications. *Council of Europe Report.*

Weller A. (2019) Transparency: Motivations and Challenges. In: Samek W., Montavon G., Vedaldi A., Hansen L., Müller KR. (eds) *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science*, vol 11700. Springer, Cham.

Weller, J.-M. (2006). Poularde farmer must be saved from remote sensing: public concern over administrative work. *Policies and Public Management*, *24*, 109-122.

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*, gvz014.

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Algorithmic decision-making and the control problem. *Minds and Machines*, 29(4), 555-578.

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

APPENDIX

Appendix A: Education Level of Participants

Table A. Descriptive statistics of participants education level

Education level	Frequency
Primary education	2
High-school	7
Post-secondary vocational education	81
Higher professional education	18
University education	6

Appendix B: Randomization Checks

Table B. Randomization checks

	Burglary	ATM Robbery	Stabbing Incident
% Female	$\chi^2(4) = 7.32$, p = .120	χ^2 (4) = 5.15, p = .272	$\chi^2(4) = 1.30$, p = .861
% Executive status	$\chi^2(4) = 4.89$, p = .299	χ^2 (4) = 2.24, p = .692	χ^2 (4) = 4.97, p = .290
% Over 5 year experience*	$\chi^2(4) = 2.14$, p = .709	$\chi^2(4) = 5.87$, p = .209	$\chi^2(4) = 5.47$, p = .242
% Higher education	$\chi^2(4) = 2.14$, p = .709	$\chi^2(4) = 3.41$, p = .492	$\chi^2(4) = 2.44$, p = .655
Average Age*	F(4, 112) = 0.63, p = .641	F(4, 112) = 0.90, p = .469	F(4, 112) = 3.76, p = .007
Knowledge about algorithms**	F(4, 112) = 1.43, p = .228	F(4, 112) = 0.45, p = .773	F(4, 112) = 0.43, p = .788
Knowledge about the HOV**	F(4, 112) = 0.89, p = .475	F(4, 112) = 0.54, p = .710	F(4, 112) = 0.72, p = .582
General trust in technology**	F(4, 112) = 1.85, p = .124	F(4, 112) = 0.39, p = .816	F(4, 112) = 2.05, p = .092

*To increase anonymity, age and experience were measured in ranges. The numbers presented are approximations based on group means. Age was measured on a 1 – 7 scale. Experience was measured on four different levels and therefore recoded to a dummy variable: 5 < year experience = 0, > 5 year experience = 1.
**Measured on a 1-7 scale

Appendix C: Validation of the Perceived Trustworthiness Scale

Table C. Principal Component Analysis and Cronbach's Alpha of the trustworthiness scale

	Burglary	ATM Robbery	Stabbing Incident
Information Correct	0.89	0.89	0.93
Advice Correct	0.87	0.9	0.94
Objective Judgment	0.92	0.9	0.89
Relevant Information	0.91	0.86	0.86
Trust the Advice	0.71	0.6	0.84
	Overall MSA = 0.88 Proportion Var 0.74 Alpha = 0.91	Overall MSA = 0.84 Proportion Var 0.70 Alpha = 0.89	Overall MSA = 0.86 Proportion Var 0.8 Alpha = 0.94

Note. Factor loadings were oblimin rotated to correct for internal correlation.

Appendix D: Descriptive Statistics of the Manipulation Checks

Table D. Descriptive statistics of the manipulation checks (SD between parentheses)

	No advice	Congruent	Incongruent	Explained congruent	Explained incongruent
Burglary					
Authenticity of the Scenario	4.88 (1.42)	4.89 (1.49)	4.84 (1.34)	5.26 (1.21)	5.48 (1.12)
Effect of explanation	4.25 (1.48)	4.21 (1.62)	4.21 (1.4)	5.04 (1.4)	4.91 (1.1)
Congruence of the advice	4.5 (1.72)	4.79 (1.69)	4.05 (1.65)	5.00 (1.48)	4.82 (1.57)
ATM Robbery					
Authenticity of the Scenario	5.17 (0.924)	5.41 (1.18)	5.44 (1.19)	5.38 (1.21)	5.12 (1.26)
Effect of explanation	4.39 (1.24)	4.71 (1.45)	4.08 (1.26)	4.66 (1.61)	4.72 (1.22)
Congruence of the advice	5.00 (1.08)	5.47 (0.943)	4.48 (1.78)	5.17 (1.34)	4.16 (1.87)
Stabbing Incident					
Authenticity of the Scenario	5.00 (1.19)	5.70 (0.865)	4.86 (1.6)	5.67 (0.637)	4.80 (1.38)
Effect of explanation	4.44 (1.29)	4.55 (1.5)	3.57 (1.69)	4.79 (1.47)	4.32 (1.68)
Congruence of the advice	4.76 (1.45)	5.40 (1.05)	3.57 (1.81)	5.42 (1.02)	4.0 (1.63)

Appendix E: Descriptive Statistics

Table E. Descriptive between and within the three scenarios

	N	% Non-PJ Choice	Trustworthiness		Certainty	
			Mean	SD	Mean	SD
Burglary						
No advice	24	50.0	4.71	1.19	48.1	24.7
Consistent advice	19	31.6	5.24	1.20	62.0	14.8
Inconsistent advice	19	42.1	4.85	1.04	47.4	22.2
Explained consistent advice	23	30.4	5.49	1.08	43.9	24.3
Explained inconsistent advice	33	51.5	5.26	0.84	56.6	25.9
ATM-robbery						
No advice	18	22.2	5.22	0.85	53.6	18.9
Consistent advice	17	5.90	5.41	0.71	60.5	23.6
Inconsistent advice	25	36.0	4.85	1.12	53.1	23.8
Explained consistent advice	29	6.90	5.02	1.36	60.0	20.9
Explained inconsistent advice	32	28.1	4.95	1.22	52.2	29.1
Stabbing incident						
No advice	25	8.00	5.09	1.02	52.0	21.4
Consistent advice	20	5.00	5.14	1.29	54.2	26.1
Inconsistent advice	28	32.1	4.55	1.51	49.4	25.8
Explained consistent advice	24	4.20	5.29	1.01	52.3	19.0
Explained inconsistent advice	25	16.0	4.66	1.42	55.1	21.1

Appendix F: The Logistic Regression Models

Table F. Logistic regression models main analyses (No advice group is the reference category)

	Burglary			ATM Robbery			Stabbing Incident			Scenarios Combined		
	Est.	S.E.	P	Est.	S.E.	P	Est.	S.E.	P	Est.	S.E.	P
Intercept	0	0.41	1	-1.25	0.57	0.027	-2.44	0.74	0.001	-1	0.28	<.001
Unexplained congruent	-0.77	0.64	0.227	-1.52	1.18	0.196	-0.5	1.26	0.691	-0.79	0.47	0.093
Unexplained incongruent	-0.32	0.62	0.607	0.68	0.7	0.336	**1.7**	**0.84**	**0.044**	0.43	0.37	0.243
Explained congruent	-0.83	0.61	0.175	-1.35	0.93	0.145	-0.69	1.26	0.582	**-0.89**	**0.44**	**0.043**
Explained incongruent	0.06	0.54	0.910	0.31	0.69	0.649	0.78	0.92	0.393	0.31	0.36	0.385
	$\chi^2(4) = 4.01$, p = .405 AIC = 166.82, BIC = 180.67			$\chi^2(4) = 11.36$, p = .023 AIC = 121.93, BIC = 135.91			$\chi^2(4) = 11.18$, p = .025 AIC = 97.34, BIC = 111.36			$\chi^2(4) = 17.95$, p = .001 AIC = 401.85, BIC = 421.30		

Appendix G: The Logistic Regression Models (Explained and Unexplained Advice Combined)

Table G. Logistic regression models exploratory analyses (No advice group is the reference category)

	Burglary			ATM robbery			Stabbing Incident			Scenarios Combined		
	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p	Est.	S.E.	p
Intercept	0.00	0.41	1.000	-1.25	0.57	.027	-2.44	0.74	.001	-1.00	0.28	<.001
Congruent advice	-0.80	0.53	.128	-1.41	0.82	.087	-0.60	1.03	.560	**-0.84**	**0.38**	**.024**
Incongruent advice	-0.08	0.49	.876	0.48	0.64	.450	1.32	0.80	.101	0.36	0.32	.258
	$\chi^2(2) = 3.57$, p = .168 AIC = 163.25, BIC = 171.57			$\chi^2(2) = 10.93$, p = .004 AIC = 118.346, BIC = 126.73			$\chi^2(2) = 9.26$, p = .010 AIC = 95.26, BIC = 103.67			$\chi^2(2) = 17.78$, p = <.001 AIC = 398.02, BIC = 409.69		

Appendix H: ANOVA Planned Contrasts Specifications

Table H. ANOVA planned contrasts

	No advice vs Advice	PJ-consists vs PJ-inconsistent	Explained vs Unexplained	PJ-inconsistent: Explanations
1. No advice (Reference)	-4	0	0	0
2. Unexplained consistent	1	1	1	0
3. Unexplained inconsistent	1	-1	1	0
4. Explained consistent	1	1	-1	1
5. Explained inconsistent	1	-1	-1	-1

Appendix I: The ANOVA on the Combined Data

Table I. Planned contrast ANOVA on combined data

	One way single ANOVA		
	Est.	S.E.	p
Intercept	5.042	0.062	<.001
No advice vs All advice types	0.013	0.032	0.671
Congruent vs Incongruent advice	**0.262**	**0.103**	**0.012**
Explained advice vs Unexplained advice	-0.06	0.069	0.384
Unexplained incongruent vs Explained Incongruent advice	-0.129	0.137	0.346
	F(4, 356) = 2.44, p = .047		

Note.

SUPPLEMENTARY MATERIAL

(See next page for the full experiment)

Welkom bij dit onderzoek!

Hartelijk dank voor je deelname aan dit onderzoek. Dit onderzoek wordt uitgevoerd door het Nationale Politie AI-Lab en de Universiteit Utrecht.

Privacyverklaring
Het doel van dit onderzoek is om inzicht te krijgen in hoe slimme algoritmen het werk van politiemedewerkers beïnvloeden. De informatie die je in de onderstaande vragenlijst opgeeft, draagt bij aan de ontwikkeling van slimme algoritmen binnen de politie en het opdoen van wetenschappelijke kennis.

In deze vragenlijst krijg je drie verzonnen situaties te zien waarin een algoritme wordt ingezet. Je wordt gevraagd hierover een aantal vragen te beantwoorden. Het invullen van de enquête duurt 8-10 minuten.

Zorg ervoor dat je het volgende begrijpt:

- **Je deelname is vrijwillig en je hebt het recht om je deelname zonder opgaaf van redenen te beëindigen.**
- **Een week na het stoppen van de dataverzameling worden de gegevens geanonimiseerd door het verwijderen van alle persoonlijke informatie die direct jou als individu kunnen identificeren.**
- **Dit betekent dat het bewaren van persoonlijk identificeerbare informatie vervalt en niemand zal kunnen identificeren welke antwoorden / gegevens van jou zijn.**
- **Dit onderzoek maakt deel uit van een afstudeerproject. De resultaten worden gepubliceerd in een masterthesis en mogelijk in een wetenschappelijk artikel. De gepubliceerde informatie zal op geen enkele wijze te herleiden zijn tot jou als persoon.**

Door op "verder" te klikken, bevestig ik dat ik het doel van dit onderzoek begrijp en begrijp hoe mijn gegevens worden verwerkt.

Voor vragen over dit onderzoek kun je contact opnemen met projectleiders dr. Stephan Grimmelikhuijsen (s.g.grimmelikhuijsen@uu.nl) of Marcel Robeer (marcel.robeer@politie.nl), of met de uitvoerder van dit onderzoek Friso Selten (friso.selten@politie.nl).

Toelichting Onderzoek

- **Deze vragenlijst wordt uitgezet onder politiemedewerkers die werkzaam zijn in verschillende functies verspreid over het land.**
- **In deze vragenlijst worden drie verschillende bedachte situaties aan je voorgelegd. Over iedere situatie worden enkele vragen gesteld.**
- **De situaties hebben niets met elkaar te maken. Het is daarom belangrijk dat je je antwoorden alleen baseert op de informatie die je in dat specifieke scenario kunt lezen.**
- **Deze vragenlijst is het meest geschikt om in te vullen op een apparaat met een groter scherm zoals een laptop, pc of tablet.**

Heb je bovenstaande tekst gelezen? Klik dan op “verder”.

Introductie Hulp Onderschepping Verdachten (HOV)

De eerste 10 minuten na de melding van een strafbaar feit zijn van cruciaal belang voor het aanhouden van een verdachte. Een centralist kan in deze eerste cruciale tien minuten niet alle eenheden op een ideale locatie positioneren. Daarom moet jij als agent eerst een eigen inschatting maken op welke plek je het best positie kunt innemen.

Om je te ondersteunen bij het maken van deze inschatting heeft de politie een computersysteem ontwikkeld: de Hulp Onderschepping Verdachten (HOV). De HOV maakt een inschatting van de vluchtroutes die verdachten kunnen gebruiken en adviseert waar jij het beste positie kan innemen om hen te onderscheppen.

Je gaat zo meedoen aan een experiment waarin je wordt gevraagd om te werken met de HOV. Over de werking van de HOV moet je weten dat:

- **Tests hebben uitgewezen dat in de meeste gevallen de HOV een betere inschatting maakt over de vluchtroute van verdachten dan een mens. We vragen je daarom het advies van de HOV serieus te nemen.**
- **Jouw eigen kennis, ervaring en intuïtie blijven echter belangrijk. De HOV geeft je een gefundeerd advies, maar je kunt ervoor kiezen dit advies niet te volgen.**

Let op: De HOV is een fictief systeem dat is gebaseerd op een systeem dat wordt ontwikkeld door de politie.

Situatie inbraak heterdaad

A 20.0% Beeld je het volgende in: jij bent aan het surveilleren in een stad in het zuidoosten van Nederland. Om 20:30 krijg je een melding via de portofoon dat er **een inbraak heterdaad** is geweest. De bewoner van een woonhuis is thuisgekomen en heeft twee inbrekers betrapt. **De verdachten zijn in noordelijke richting gevlucht op een scooter met een Duits kenteken.**

Jij bevindt je in de omgeving van deze inbraak en wordt gevraagd om te helpen met het met het met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze inbraak.
- **De HOV heeft berekend dat de kans dat de verdachten vluchten langs locatie A of langs Locatie B even groot is.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.

B 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een stad in het zuidoosten van Nederland. Om 20:30 krijg je een melding via de portofoon dat er **een inbraak heterdaad** is geweest. De bewoner van een woonhuis is thuisgekomen en heeft twee inbrekers betrapt. **De verdachten zijn in noordelijke richting gevlucht op een scooter met een Duits kenteken.**

Jij bevindt je in de omgeving van deze inbraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze inbraak.
- **De HOV heeft berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.

C 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een stad in het zuidoosten van Nederland. Om 20:30 krijg je een melding via de portofoon dat er **een inbraak heterdaad** is geweest. De bewoner van een woonhuis is thuisgekomen en heeft twee inbrekers betrapt. **De verdachten zijn in noordelijke richting gevlucht op een scooter met een Duits kenteken.**

Jij bevindt je in de omgeving van deze inbraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze inbraak.
- **De HOV heeft berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.

D 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een stad in het zuidoosten van Nederland. Om 20:30 krijg je een melding via de portofoon dat er **een inbraak heterdaad** is geweest. De bewoner van een woonhuis is thuisgekomen en heeft twee inbrekers betrapt. **De verdachten zijn in noordelijke richting gevlucht op een scooter met een Duits kenteken.**

Jij bevindt je in de omgeving van deze inbraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze inbraak.
- De HOV heeft **Locatie A** overwogen omdat dit de snelste route richting Duitsland is. De HOV heeft **Locatie B** overwogen omdat verdachten van inbraken vaak via rustige wegen vluchten.
- **De HOV heeft op basis van deze informatie berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.

E 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een stad in het zuidoosten van Nederland. Om 20:30 krijg je een melding via de portofoon dat er **een inbraak heterdaad** is geweest. De bewoner van een woonhuis is thuisgekomen en heeft twee inbrekers betrapt. **De verdachten zijn in noordelijke richting gevlucht op een scooter met een Duits kenteken.**

Jij bevindt je in de omgeving van deze inbraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze inbraak.
- De HOV heeft **Locatie A** overwogen omdat verdachten van inbraken vaak via rustige wegen vluchten. De HOV heeft **Locatie B** overwogen omdat dit de snelste route richting Duitsland is.
- **De HOV heeft op basis van deze informatie berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.

* 1. Op basis van het advies van de HOV en jouw eigen inschatting van de situatie, welke locatie kies je dan om positie in te nemen om de verdachten van de inbraak te onderscheppen?

- ◯ Locatie A
- ◯ Locatie B

POLITIE

Situatie inbraak heterdaad

* 2. Je hebt de keuze gemaakt om naar [V1] te gaan. Hoe zeker ben je er van dat de verdachten daadwerkelijk langs deze route gevlucht zijn?

0% (helemaal niet zeker) 100% (helemaal zeker)

* 3. Op basis van de informatie die je zojuist hebt gelezen vragen we je een inschatting te maken over de betrouwbaarheid van het advies dat de HOV je heeft gegeven over de vluchtroute van de verdachten van de inbraak.

Ik vertrouw erop dat de HOV bij het opstellen van dit advies...

	1: compleet mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee oneens, niet mee eens	5: een beetje mee eens	6: mee eens	7: compleet mee eens
…de juiste informatie heeft gebruikt.	○	○	○	○	○	○	○
…een correct aanbeveling heeft gegeven.	○	○	○	○	○	○	○
…de situatie objectief heeft beoordeeld.	○	○	○	○	○	○	○
…alle relevante informatie heeft afgewogen.	○	○	○	○	○	○	○

* 4. Ik vind het advies wat de HOV mij gaf over de vluchtroute van de verdachten van de inbraak:

1: compleet onbetrouwbaar	2: onbetrouwbaar	3: een beetje onbetrouwbaar	4: niet onbetrouwbaar, niet betrouwbaar	5: een beetje betrouwbaar	6: betrouwbaar	7: compleet betrouwbaar
○	○	○	○	○	○	○

Situatie inbraak heterdaad

* 5. Geef aan in hoeverre de volgende stellingen van toepassing zijn op de inbraak heterdaad en de inzet van de HOV in deze situatie.

	1: helemaal mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee eens, niet mee oneens	5: een beetje mee eens	6: mee eens	7: helemaal mee eens
Het advies van de HOV was gedetailleerd.	○	○	○	○	○	○	○
De HOV gaf duidelijk aan wat de beste locatie was waar ik me moest opstellen.	○	○	○	○	○	○	○
Het advies sluit goed aan bij mijn eigen inschatting van de vluchtroute van de verdachten.	○	○	○	○	○	○	○
Ik kon mij inleven in de aan mij voorgelegde situatie.	○	○	○	○	○	○	○

POLITIE

Situatie 2/3

Op de volgende pagina wordt een nieuwe situatie gepresenteerd. Het is hierbij van belang dat je dit als een compleet nieuwe situatie ziet. Bij het beantwoorden van de vragen in de volgende situatie is het niet nodig om rekening te houden met de antwoorden die je in de vorige situatie hebt gegeven.

Heb je alle bovenstaande tekst gelezen? Klik dan op “verder”.

Situatie plofkraak

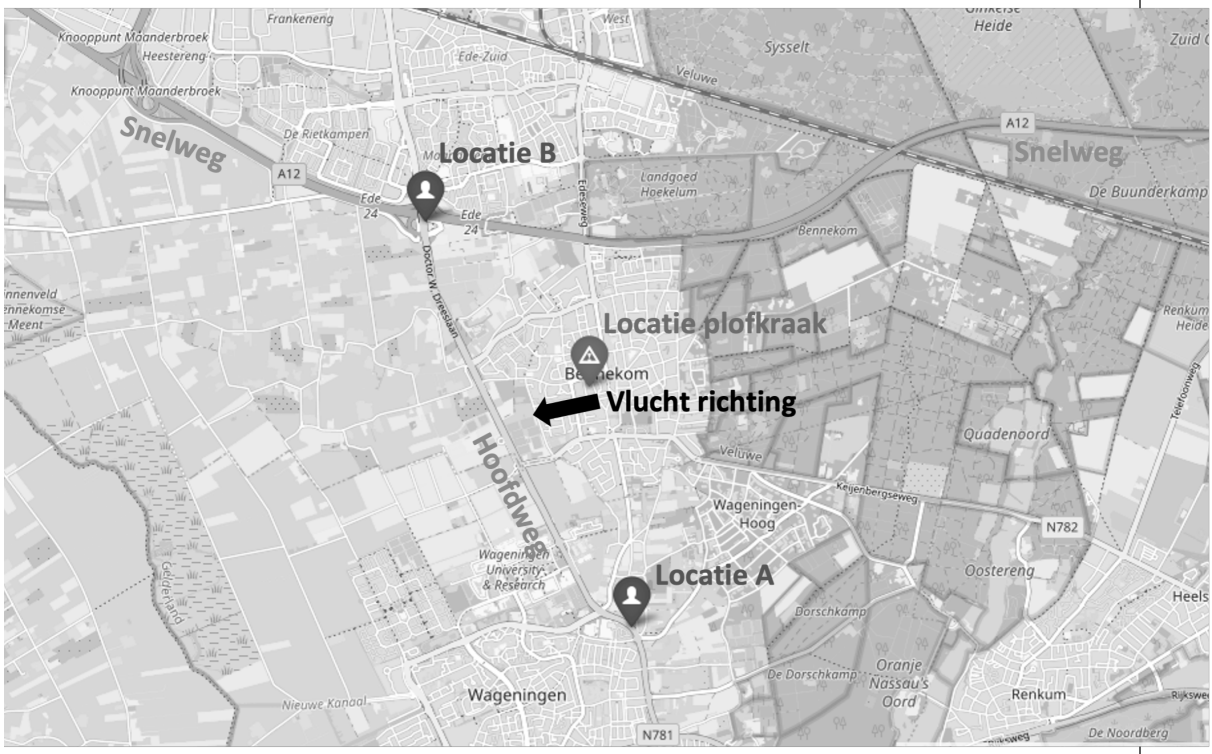
A 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een regio in het midden van Nederland. Om 18:30 krijg je een melding via de portofoon dat er **een plofkraak** is geweest. De getuige die het incident meldt heeft direct 112 gebeld. Via de portofoon hoor je dat een getuige heeft gezien dat er **twee verdachten zijn, die zijn gevlucht in westelijke richting in een grijze sportwagen.**

Jij bevindt je in de omgeving van deze plofkraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze plofkraak.
- **De HOV heeft berekend dat de kans dat de verdachten vluchten langs locatie A of langs Locatie B even groot is.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



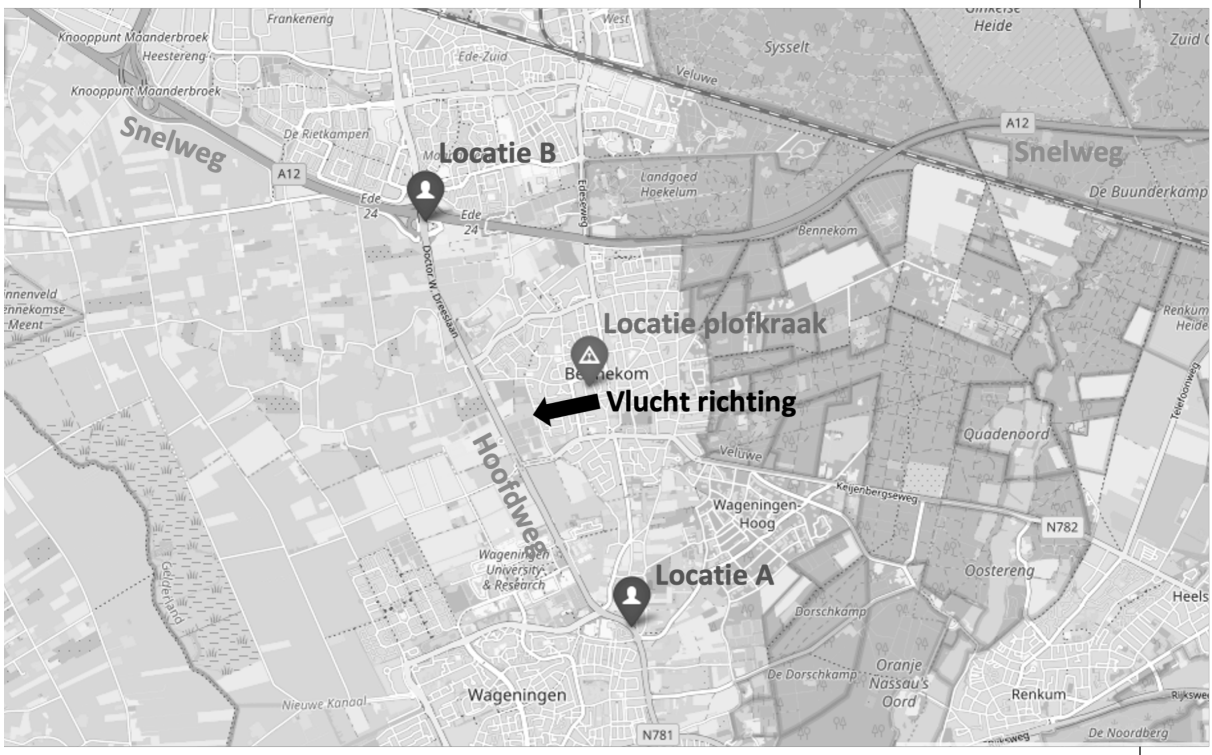
B 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een regio in het midden van Nederland. Om 18:30 krijg je een melding via de portofoon dat er **een plofkraak** is geweest. De getuige die het incident meldt heeft direct 112 gebeld. Via de portofoon hoor je dat een getuige heeft gezien dat er **twee verdachten zijn, die zijn gevlucht in westelijke richting in een grijze sportwagen.**

Jij bevindt je in de omgeving van deze plofkraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze plofkraak.
- **De HOV heeft berekend dat je van deze twee locaties het beste naar locatie B kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



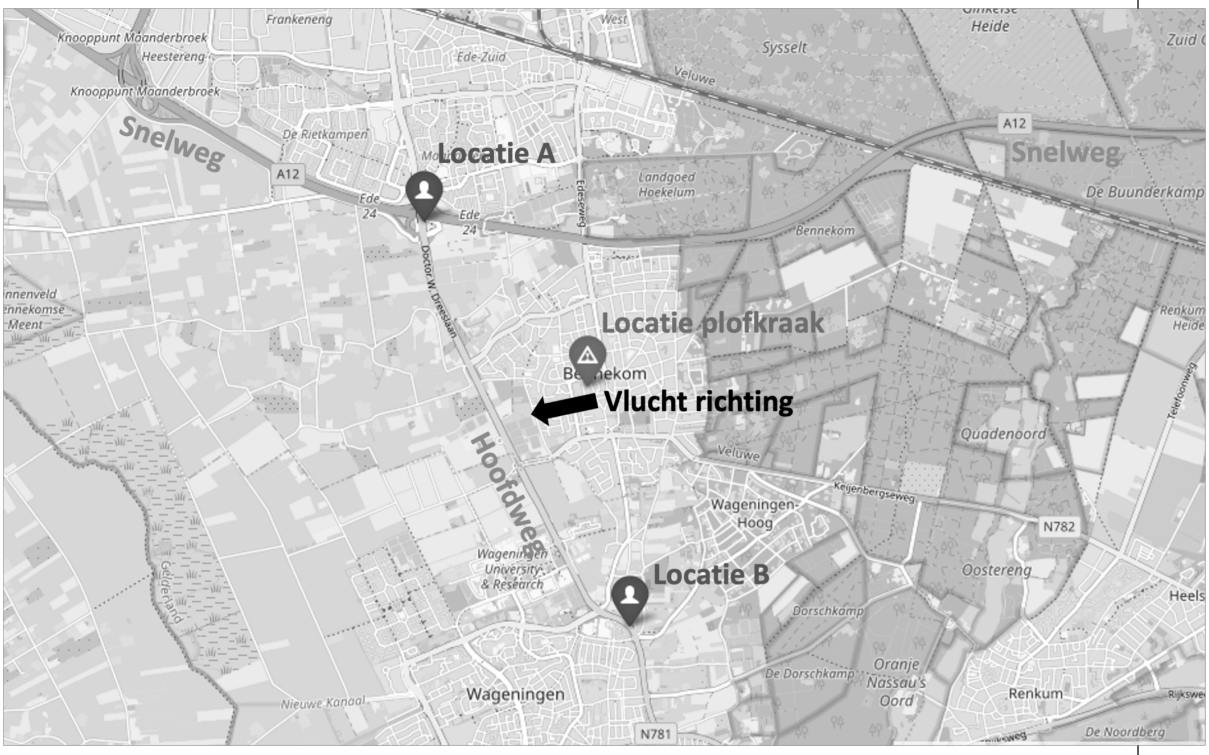
C 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een regio in het midden van Nederland. Om 18:30 krijg je een melding via de portofoon dat er **een plofkraak** is geweest. De getuige die het incident meldt heeft direct 112 gebeld. Via de portofoon hoor je dat een getuige heeft gezien dat er **twee verdachten zijn, die zijn gevlucht in westelijke richting in een grijze sportwagen.**

Jij bevindt je in de omgeving van deze plofkraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze plofkraak.
- **De HOV heeft berekend dat je van deze twee locaties het beste naar locatie B kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



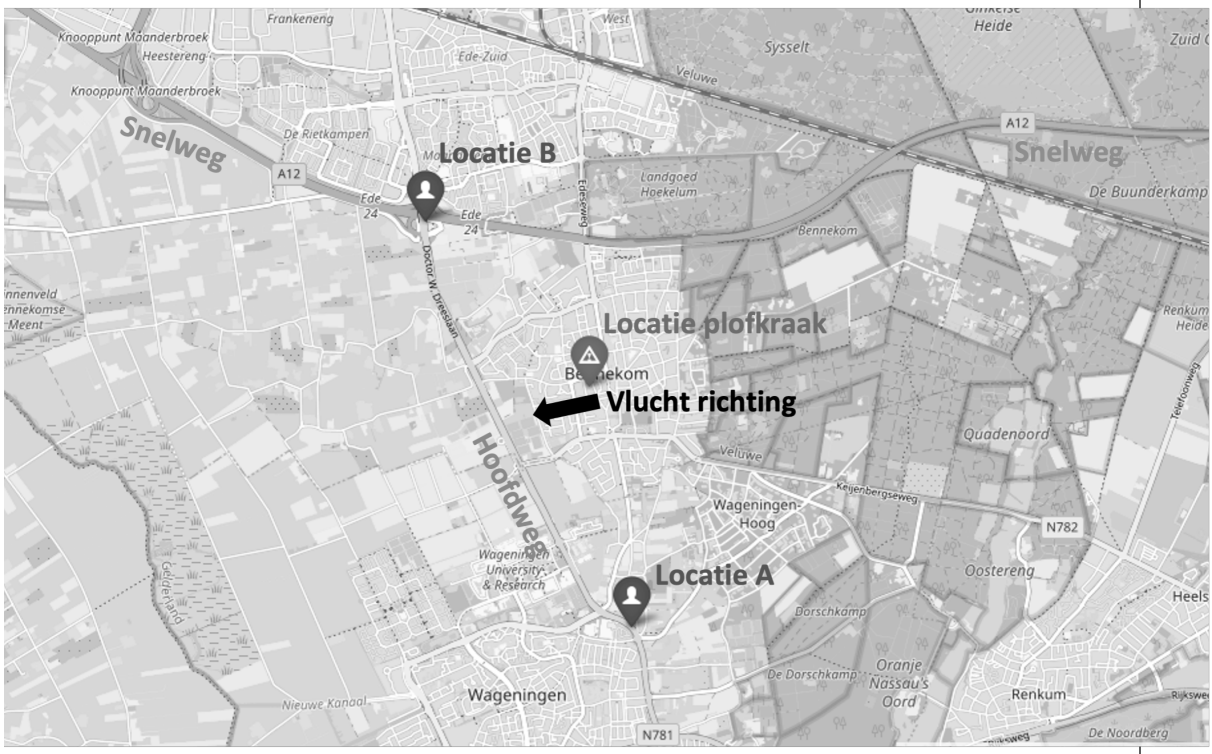
D 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een regio in het midden van Nederland. Om 18:30 krijg je een melding via de portofoon dat er **een plofkraak** is geweest. De getuige die het incident meldt heeft direct 112 gebeld. Via de portofoon hoor je dat een getuige heeft gezien dat er **twee verdachten zijn, die zijn gevlucht in westelijke richting in een grijze sportwagen.**

Jij bevindt je in de omgeving van deze plofkraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze plofkraak.
- De HOV heeft **Locatie A** overwogen omdat de verdachten dan niet via deze route de aangrenzende stad in kunnen vluchten. De HOV heeft **Locatie B** overwogen omdat verdachten van plofkraken vaak via de snelweg vluchten.
- **De HOV heeft op basis van deze informatie berekend dat je van deze twee locaties het beste naar Locatie B kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



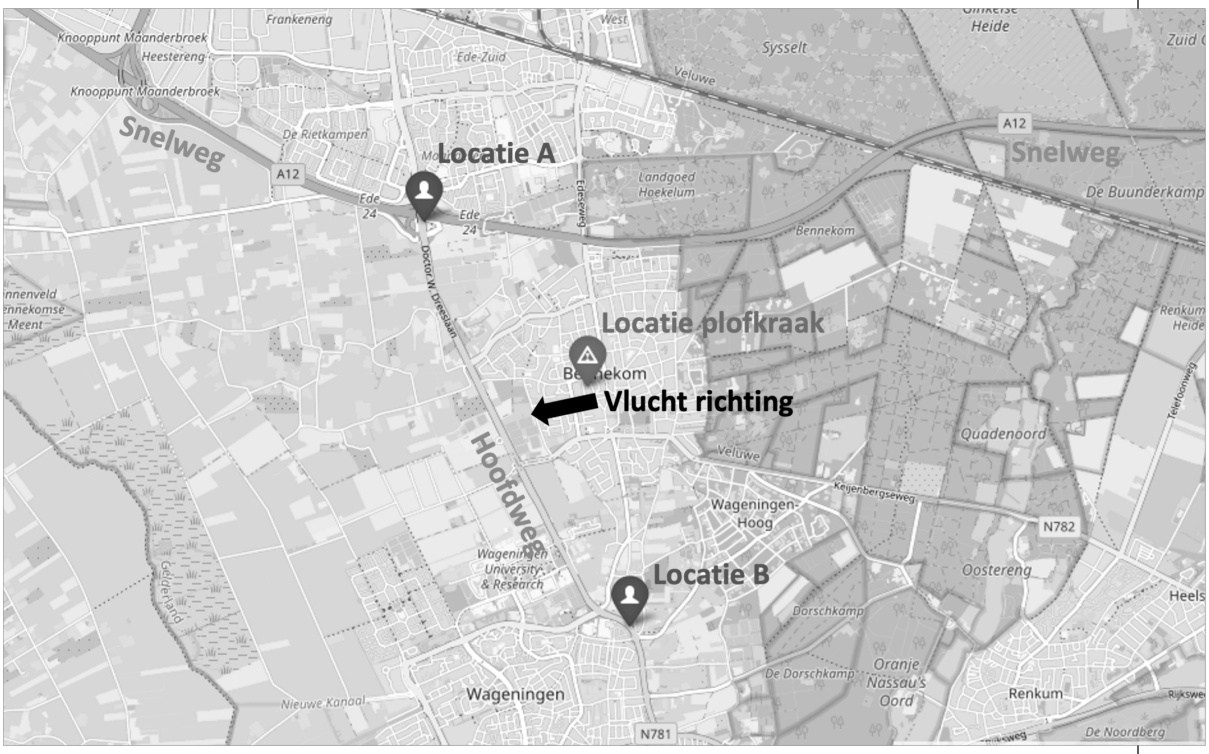
E 20.0% Beeld je het volgende in: Jij bent aan het surveilleren in een regio in het midden van Nederland. Om 18:30 krijg je een melding via de portofoon dat er **een plofkraak** is geweest. De getuige die het incident meldt heeft direct 112 gebeld. Via de portofoon hoor je dat een getuige heeft gezien dat er **twee verdachten zijn, die zijn gevlucht in westelijke richting in een grijze sportwagen.**

Jij bevindt je in de omgeving van deze plofkraak en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachten aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachten van deze plofkraak.
- De HOV heeft **Locatie A** overwogen omdat verdachten van plofkraken vaak via de snelweg vluchten. De HOV heeft **Locatie B** overwogen omdat de verdachten dan niet via deze route de aangrenzende stad in kunnen vluchten.
- **De HOV heeft op basis van deze informatie berekend dat je van deze twee locaties het beste naar Locatie B kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



* 6. Op basis van het advies van de HOV en jouw eigen inschatting van de situatie, welke locatie kies je dan om positie in te nemen om de verdachten van de plofkraak te onderscheppen?

- ◯ Locatie A
- ◯ Locatie B

Situatie plofkraak

* 7. Je hebt de keuze gemaakt om naar [V6] te gaan. Hoe zeker ben je er van dat de verdachten daadwerkelijk langs deze route gevlucht zijn?

0% (helemaal niet zeker) 100% (helemaal zeker)

* 8. Op basis van de informatie die je zojuist hebt gelezen vragen we je een inschatting te maken over de betrouwbaarheid van het advies dat de HOV je heeft gegeven over de vluchtroute van de verdachten van de plofkraak.

Ik vertrouw erop dat de HOV bij het opstellen van dit advies...

	1: compleet mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee oneens, niet mee eens	5: een beetje mee eens	6: mee eens	7: compleet mee eens
…de juiste informatie heeft gebruikt.							
…een correct aanbeveling heeft gegeven.							
…de situatie objectief heeft beoordeeld.							
…alle relevante informatie heeft afgewogen.							

* 9. Ik vind het advies wat de HOV mij gaf over de vluchtroute van de verdachten van de plofkraak:

1: compleet onbetrouwbaar	2: onbetrouwbaar	3: een beetje onbetrouwbaar	4: niet onbetrouwbaar, niet betrouwbaar	5: een beetje betrouwbaar	6: betrouwbaar	7: compleet betrouwbaar

Situatie plofkraak

* 10. Geef aan in hoeverre de volgende stellingen van toepassing zijn op de plofkraak en de inzet van de HOV in deze situatie.

	1: helemaal mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee eens, niet mee oneens	5: een beetje mee eens	6: mee eens	7: helemaal mee eens
Het advies van de HOV was gedetailleerd.	○	○	○	○	○	○	○
De HOV gaf duidelijk aan wat de beste locatie was waar ik me moest opstellen.	○	○	○	○	○	○	○
Het advies sluit goed aan bij mijn eigen inschatting van de vluchtroute van de verdachten.	○	○	○	○	○	○	○
Ik kon mij inleven in de aan mij voorgelegde situatie.	○	○	○	○	○	○	○

POLITIE

Situatie 3/3

Op de volgende pagina wordt de laatste situatie gepresenteerd. Het is hierbij van belang dat je deze wederom als een compleet nieuwe situatie ziet. Bij het beantwoorden van de vragen in de volgende situatie is het niet nodig om rekening te houden met de antwoorden die je in de vorige situatie hebt gegeven.

Heb je alle bovenstaande tekst gelezen? Klik dan op “verder”.

Situatie steekincident

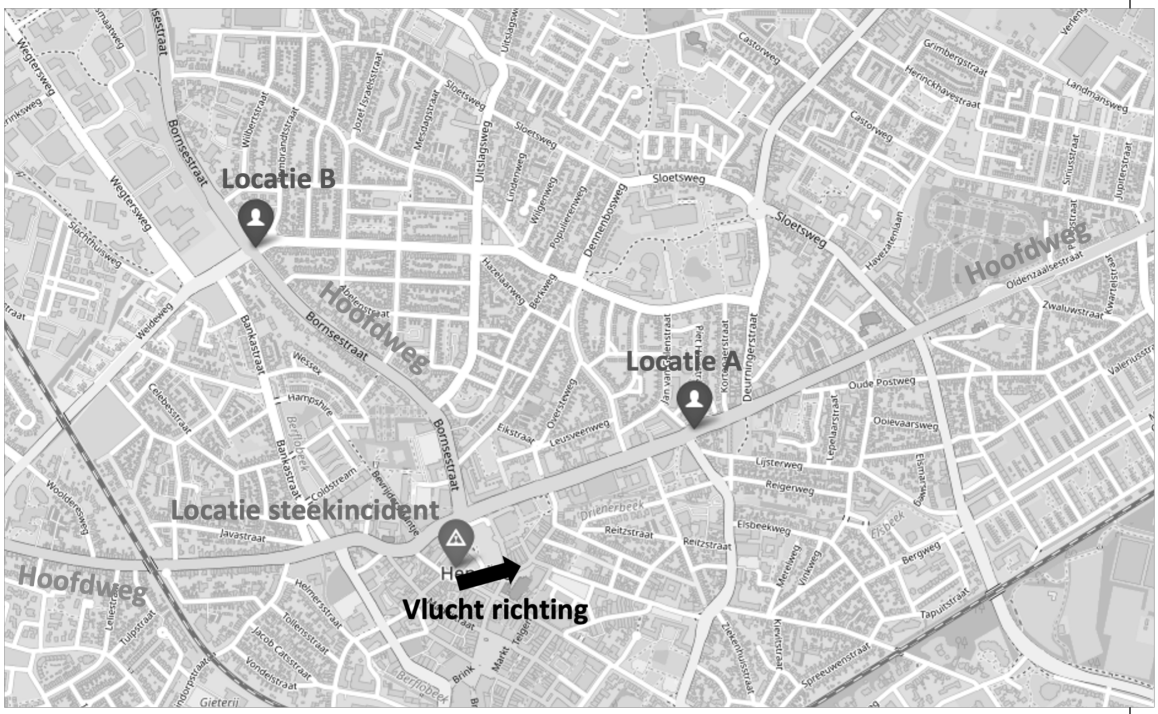
A 20.0% Beeld je het volgende in: Jij hebt nachtdienst in een middelgrote stad in Nederland. Om 03:30 krijg je een melding dat er **een steekincident** heeft plaatsgevonden. Via de portofoon hoor je dat cameratoezicht heeft gezien dat tijdens een ruzie iemand met een mes is gestoken. **De verdachte is na het incident in een rode auto in oostelijke richting weggereden**. Cameratoezicht heeft geen zicht meer op de verdachte.

Jij bevindt je in de omgeving van dit steekincident en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachte aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachte van dit steekincident.
- **De HOV heeft berekend dat de kans dat de verdachte vlucht langs Locatie A of langs Locatie B even groot is.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



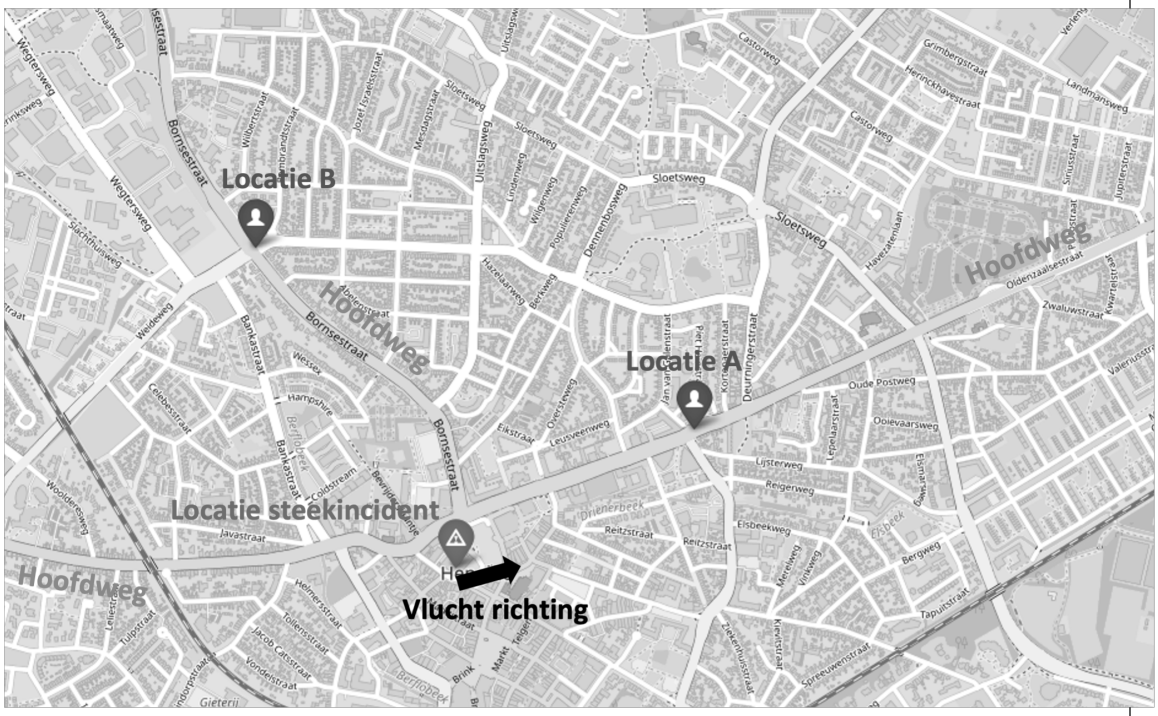
B 20.0% Beeld je het volgende in: Jij hebt nachtdienst in een middelgrote stad in Nederland. Om 03:30 krijg je een melding dat er **een steekincident** heeft plaatsgevonden. Via de portofoon hoor je dat cameratoezicht heeft gezien dat tijdens een ruzie iemand met een mes is gestoken. **De verdachte is na het incident in een rode auto in oostelijke richting weggereden**. Cameratoezicht heeft geen zicht meer op de verdachte.

Jij bevindt je in de omgeving van dit steekincident en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachte aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachte van dit steekincident.
- **De HOV heeft berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



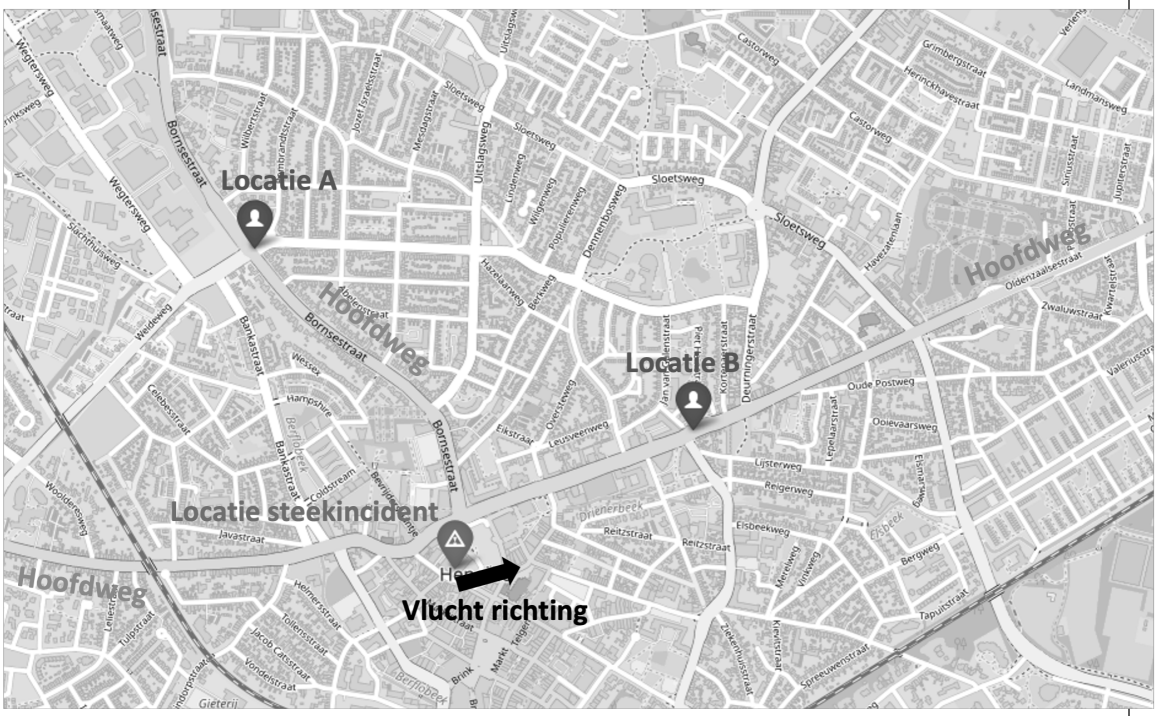
C 20.0% Beeld je het volgende in: Jij hebt nachtdienst in een middelgrote stad in Nederland. Om 03:30 krijg je een melding dat er **een steekincident** heeft plaatsgevonden. Via de portofoon hoor je dat cameratoezicht heeft gezien dat tijdens een ruzie iemand met een mes is gestoken. **De verdachte is na het incident in een rode auto in oostelijke richting weggereden**. Cameratoezicht heeft geen zicht meer op de verdachte.

Jij bevindt je in de omgeving van dit steekincident en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachte aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachte van dit steekincident.
- **De HOV heeft berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



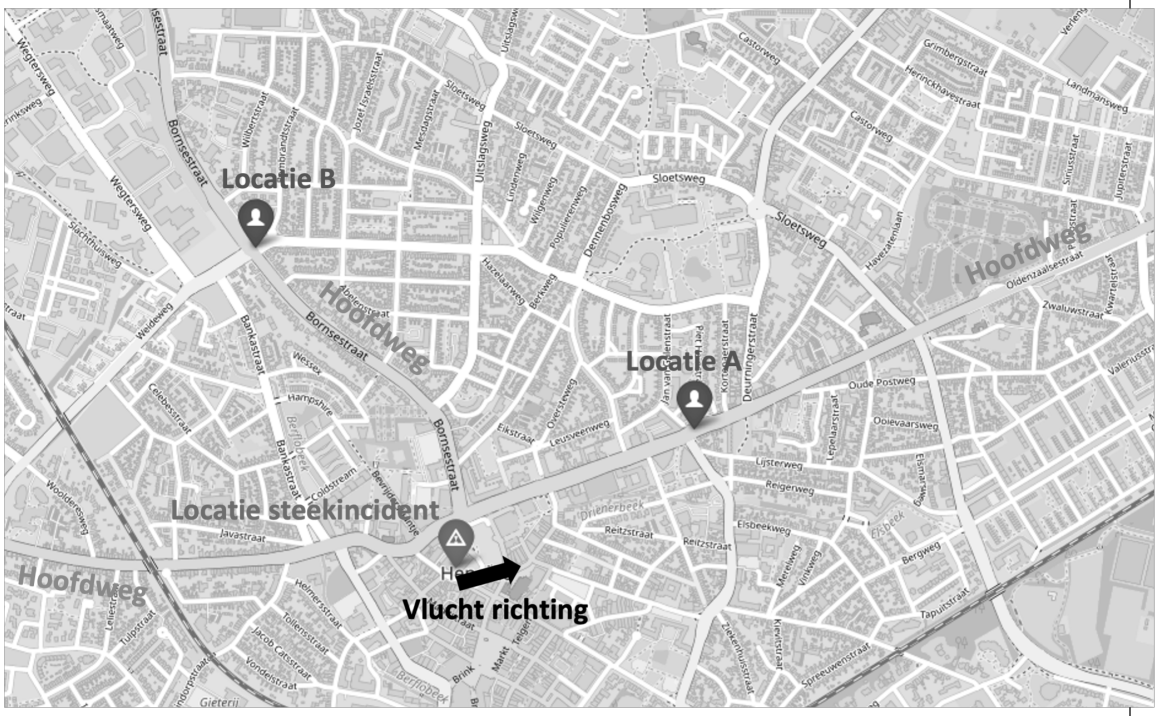
D 20.0% Beeld je het volgende in: Jij hebt nachtdienst in een middelgrote stad in Nederland. Om 03:30 krijg je een melding dat er **een steekincident** heeft plaatsgevonden. Via de portofoon hoor je dat cameratoezicht heeft gezien dat tijdens een ruzie iemand met een mes is gestoken. **De verdachte is na het incident in een rode auto in oostelijke richting weggereden**. Cameratoezicht heeft geen zicht meer op de verdachte.

Jij bevindt je in de omgeving van dit steekincident en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachte aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachte van dit steekincident.
- De HOV heeft **Locatie A** overwogen omdat dit een centraal punt langs een grote weg in de vluchtrichting (oostelijk) van de verdachte is. De HOV heeft **Locatie B** overwogen omdat dit een centrale locatie is waar veel mogelijke vluchtroutes samenkomen.
- **De HOV heeft op basis van deze informatie berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



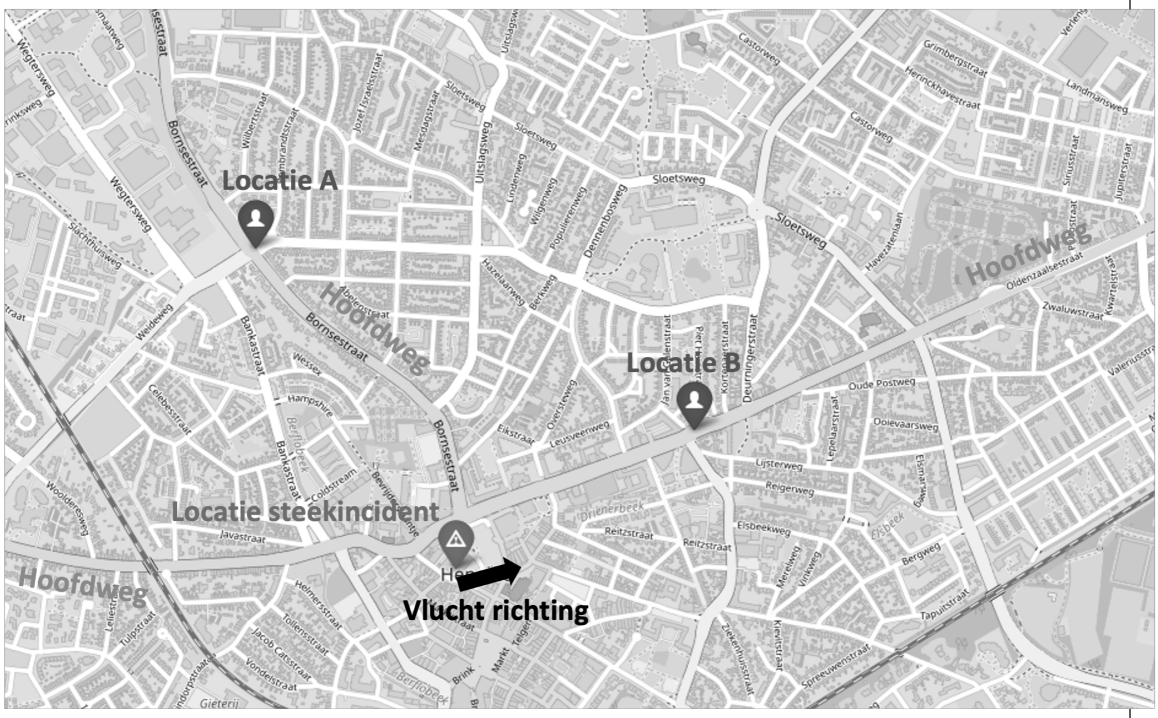
E 20.0% Beeld je het volgende in: Jij hebt nachtdienst in een middelgrote stad in Nederland. Om 03:30 krijg je een melding dat er **een steekincident** heeft plaatsgevonden. Via de portofoon hoor je dat cameratoezicht heeft gezien dat tijdens een ruzie iemand met een mes is gestoken. **De verdachte is na het incident in een rode auto in oostelijke richting weggereden.** Cameratoezicht heeft geen zicht meer op de verdachte.

Jij bevindt je in de omgeving van dit steekincident en wordt gevraagd om te helpen met het dichtzetten van vluchtroutes. Je hebt van de meldkamer doorgekregen dat er weinig andere eenheden in de buurt van het incident zijn.

Om de kans om de verdachte aan te houden te vergroten wordt de Hulp Onderschepping Verdachten (HOV) ingeschakeld.

- Via een app krijg je de melding dat de HOV heeft bepaald dat **Locatie A en Locatie B** mogelijke vluchtroutes zijn voor de verdachte van dit steekincident.
- De HOV heeft **Locatie A** overwogen omdat dit een centrale locatie is waar veel mogelijke vluchtroutes samenkomen. De HOV heeft **Locatie B** overwogen omdat dit een centraal punt langs een grote weg in de vluchtrichting (oostelijk) van de verdachte is.
- **De HOV heeft op basis van deze informatie berekend dat je van deze twee locaties het beste naar Locatie A kunt gaan.**

Onderstaande kaart toont de verschillende locaties die in deze situatie worden genoemd. We vragen je deze kaart goed te bestuderen om zo een duidelijk beeld van de situatie te krijgen.



* 11. Op basis van het advies van de HOV en jouw eigen inschatting van de situatie, welke locatie kies je dan om positie in te nemen om de verdachte van het steekincident te onderscheppen?

- ◯ Locatie A
- ◯ Locatie B

Situatie steekincident

* 12. Je hebt de keuze gemaakt om naar [V11] te gaan. Hoe zeker ben je er van dat de verdachte daadwerkelijk langs deze route gevlucht is?

0% (helemaal niet zeker) 100% (helemaal zeker)

* 13. Op basis van de informatie die je zojuist hebt gelezen vragen we je een inschatting te maken over de betrouwbaarheid van het advies dat de HOV je heeft gegeven over de vluchtroute van de verdachte van het steekincident.

Ik vertrouw erop dat de HOV bij het opstellen van dit advies...

	1: compleet mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee oneens, niet mee eens	5: een beetje mee eens	6: mee eens	7: compleet mee eens
…de juiste informatie heeft gebruikt.							
…een correct aanbeveling heeft gegeven.							
…de situatie objectief heeft beoordeeld.							
…alle relevante informatie heeft afgewogen.							

* 14. Ik vind het advies wat de HOV mij gaf over de vluchtroute van de verdachte van het steekincident:

1: compleet onbetrouwbaar	2: onbetrouwbaar	3: een beetje onbetrouwbaar	4: niet onbetrouwbaar, niet betrouwbaar	5: een beetje betrouwbaar	6: betrouwbaar	7: compleet betrouwbaar

Situatie steekincident

* 15. Geef aan in hoeverre de volgende stellingen van toepassing zijn op het steekincident en de inzet van de HOV in deze situatie.

	1: helemaal mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee eens, niet mee oneens	5: een beetje mee eens	6: mee eens	7: helemaal mee eens
Het advies van de HOV was gedetailleerd.	○	○	○	○	○	○	○
De HOV gaf duidelijk aan wat de beste locatie was waar ik me moest opstellen.	○	○	○	○	○	○	○
Het advies sluit goed aan bij mijn eigen inschatting van de vluchtroute van de verdachte.	○	○	○	○	○	○	○
Ik kon mij inleven in de aan mij voorgelegde situatie.	○	○	○	○	○	○	○

Dit was de laatste situatie. Er volgen nu nog zeven korte vragen over jouw achtergrond en je werkzaamheden bij de politie.

POLITIE

* 16. Wat is je geslacht?

- Vrouw
- Man
- Anders
- Wil ik niet zeggen

* 17. Welke leeftijd heb je?

* 18. Geef aan in hoeverre de volgende stellingen op jou van toepassing zijn.

	1: helemaal mee oneens	2: mee oneens	3: een beetje mee oneens	4: niet mee eens, niet mee oneens	5: een beetje mee eens	6: mee eens	7: Helemaal mee eens
Ik heb veel kennis over de werking van slimme algoritmen.							
Ik heb veel vertrouwen in technologie.							
In mijn werk bij de politie kom ik vaak in aanraking met algoritmen die lijken op de HOV.							

* 19. In welke politieregio ben je het meeste werkzaam?

- Noord-Nederland
- Oost-Nederland
- Midden-Nederland
- Noord-Holland
- Amsterdam
- Den Haag
- Rotterdam
- Zeeland - West-Brabant
- Oost-Brabant
- Limburg
- Landelijke eenheid
- Wil ik niet zeggen

* 20. Hoe lang ben je al bij de politie werkzaam?

- Minder dan 2 jaar
- 2-5 jaar
- 6-10 jaar
- Meer dan 10 jaar
- Wil ik niet zeggen

* 21. Heb je een executieve status?

- ◯ Ja
- ◯ Nee
- ◯ Nee, maar ik heb wel een executieve status gehad
- ◯ Nee, maar ik ben in opleiding voor een executieve status
- ◯ Wil ik niet zeggen

* 22. Wat is de hoogste opleiding die je hebt voltooid?

POLITIE

Hartelijk bedankt voor je deelname!

In dit onderzoek is getest wat de invloed van het uitleggen van voorspelling van een algoritme op het vertrouwen in algoritmische voorspellingen. Van ieder van de drie situaties (de inbraak, de plofkraak en het steekincident) zijn er vijf verschillende varianten die verschillende type voorspellingen bevatten. In dit onderzoek is voor ieder van deze situaties een van deze vijf varianten aan jou gepresenteerd.

De situaties die zijn gebruikt in dit onderzoek zijn fictief en ook de HOV is een fictief systeem. Ondanks het fictieve karakter van de situaties in deze enquête draagt jouw deelname direct bij aan de ontwikkeling van slimme algoritmen binnen de politie en wetenschappelijke kennis over dit onderwerp. Meer informatie over de ontwikkeling van slimme algoritme binnen de politie kun je vinden op de website van het Nationale Politie AI-lab (https://icai.ai/police-lab-ai/).

Dit onderzoek wordt uitgevoerd door het Nationale Politie AI-lab en de Universiteit Utrecht. Voor vragen over dit onderzoek kun je contact opnemen met projectleiders dr. Stephan Grimmelikhuijsen (s.g.grimmelikhuijsen@uu.nl) of Marcel Robeer (marcel.robeer@politie.nl), of met de uitvoerder van dit onderzoek Friso Selten (friso.selten@politie.nl).