

Handling Heuristics

How behavioral sciences can support
decision-making in the public sector



Rosanna Nagtegaal

Handling Heuristics:

How behavioral sciences can support decision-making in the public sector

Rosanna Nagtegaal

Doctoral committee

Prof. Dr. Don Moynihan (Georgetown University, United States of America)

Prof. Dr. Denise de Ridder (Utrecht University, the Netherlands)

Prof. Dr. Sandra van Thiel (Erasmus University Rotterdam, the Netherlands)

Prof. Dr. Eva Thomann (University of Konstanz, Germany)

Prof. Dr. Barbara Vis (Utrecht University, the Netherlands)

Handling Heuristics: How Behavioral Sciences can Support Decision-Making in the Public Sector

ISBN: 978-94-6458-010-5

Cover design: Publiss | www.publiss.nl

Internal design: Publiss | www.publiss.nl

English editing: Giles Stacey and Bryan Spooner

Print: Ridderprint | www.ridderprint.nl

Copyright © 2022 Rosanna Nagtegaal

All rights reserved. No part of this thesis may be reproduced, stored or transmitted in any way or by any means without the prior permission of the author, or when applicable, of the publishers of the scientific papers.

Handling Heuristics

How behavioral sciences can support decision-making
in the public sector

Heuristieken hanteren

Hoe gedragswetenschap besluitvorming in de publieke sector
kan ondersteunen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 18 maart 2022 des middags te 4.15 uur

door

Rosanna Nagtegaal

geboren op 31 juli 1990
te Utrecht

PROMOTOREN:

Prof. dr. L.G. Tummers

Prof. dr. M. Noordegraaf

Prof. dr. V.J.J.M. Bekkers

CONTENTS

Chapter 1:	Introduction	7
Chapter 2:	Nudging healthcare professionals towards evidence-based medicine: A systematic scoping review	29
Chapter 3:	Designing to debias: Measuring and reducing public managers' anchoring bias	49
Chapter 4:	Don't speak: exploring how personal and contextual factors affect politicians' participation in online deliberations	69
Chapter 5:	The impact of using algorithms for managerial decisions on public employees' procedural justice	87
Chapter 6:	Conclusions and implications	105
	Summary	123
	Samenvatting in het Nederlands	137
	References	151
	Appendices	175
	Acknowledgements	191
	About the author	197



CHAPTER 1

Introduction

Laila and José work as general practitioners (GPs). José has worked as an independent GP in a rural region for many years. Laila just started in a large city practice, and sometimes collaborates with other GPs. They both have a lot of patients given the general practitioner shortages in their regions. To keep up to date, they regularly follow courses. Last month, they both for instance followed an online educational module about the prescription of antibiotics. In this module, Laila and José were reminded that healthcare professionals prescribe too many antibiotics for respiratory tract infections, such as a common cold. Taking antibiotics for these types of infections does not heal patients. In fact, taking too many antibiotics can make patients resistant to antibiotics. If this happens, antibiotics will no longer be as effective anymore and might be unable to heal as many people.

After the educational module, Laila received information about alternatives to antibiotics. Although she found this interesting, this information did not change her prescription habits significantly. At the same time, José received an e-mail about how many inappropriate antibiotics he had prescribed and whether he was a top performer or not. To his surprise, he saw that he was prescribing many more inappropriate antibiotics than some of his peers. Subsequently, he decided to prescribe fewer antibiotics. Later, one patient came back to José's practice with a bacterial infection. It remained unclear if this patient could have benefitted from antibiotics.

The example, based on Meeker et al. (2016), shows that decisions in the public sector happen within choice environments that can be adapted (Münscher et al., 2016). In the example, these choice environments are the healthcare industry, the busy rural and urban practices in which Laila and José work, as well as the digital environments in which they receive information related to their decisions. Choice environments have effects on decision-making because people make decisions by using *heuristics*. Heuristics are shortcuts for making decisions, that allow people to make decisions easily and quickly (Kahneman et al., 1982; Tversky & Kahneman, 1974). Heuristics are often very helpful, but they can also lead to suboptimal decision-making. In these cases, they are called biases.

An example of a bias is anchoring. This is the tendency to use initial values when estimating unknown quantities (Tversky & Kahneman, 1974). These initial values are sometimes helpful. When asked, for instance, to estimate the boiling point of water on Mount Everest, people might start from what they know about the boiling point of water at sea level (100 °C) and adjust their response from there (Lieder et al., 2017). Nevertheless, anchoring can also create biased decision-making, as the numbers presented are sometimes irrelevant. This for example happens in property appraisals. During property appraisals, almost exactly similar houses can be judged differently. Ünveren & Baycar (2019) found that this happened because appraisals are subconsciously adjusted to house numbers.

We can use heuristics to influence decision-making. These insights have been used for many years in commercial organizations. Consider the pricing of products at 9.99, instead of at 10 euros (Thomas & Morwitz, 2005). These insights have gained attention recently in public administration research and practice. Governments for instance started nudging

citizens to achieve policy goals (Dolan et al., 2012). An example is changing organ donation systems from being opt-in to opt-out (Johnson & Goldstein, 2003). This development has been described by the behavioral public administration (BPA) stream, which explicitly integrates psychology with public administration (e.g., Grimmelikhuijsen et al., 2017). So far, the BPA stream has mainly focused on the application of heuristics and biases to citizens. Scholars and practitioners have, however, started to apply insights on heuristics to studying decision-making by public sector workers (Sunstein & Thaler 2008; Soman & Yeung 2021; Hallsworth et al. 2018). Scholars and practitioners have also started to look at applications of behavioral sciences beyond heuristics (Ewert, 2020; Gofen et al., 2021). In other words, we have started to explore how we can *handle heuristics* conceptually as well as empirically to support decision-making within the public sector.

As we see in our example, the implications of using behavioral sciences to intervene in public sector decision-making can be large. For instance, in the short term, general practitioners' decision-making affects if a person heals or not. In the example, one person came back to José with a bacterial infection that might have been avoided. In the longer term, fewer antibiotic prescriptions can keep antibiotics effective for future patients. Although this example is about service provision in healthcare, the same is true for decisions in other domains in the public sector. Politicians for example make daily decisions that affect who gets what how and when as described by Lasswell (1936). Public managers balance organizational and employees' needs, and through their employees they influence citizens. This is why studying decision-making has always been at the heart of public administration (Simon, 1997).

The example also shows that simple, cheap interventions based on heuristics have the potential to support public sector workers' decision-making. Laila received information about alternatives to antibiotics, while José was subject to a social norm intervention, which was a simple e-mail with information about inappropriate antibiotics prescribed and about being a top-performer or not. The social norm intervention had an impact on the amount of antibiotics prescribed, while the suggested alternatives intervention did not (Meeker et al., 2016). Thus, using knowledge about heuristics and biases has the potential to offer simple and cheap solutions to support public sector workers' decision-making, which is why scholars have turned their attention to its potential in the public sector (Hallsworth et al., 2018).

Moving our attention from consumers and citizens to public sector workers comes with unique challenges. These challenges concern the specific *contexts* that public sector workers operate in and the *individual characteristics* that public sector workers have. Public organizations are shaped by *institutional* and *situational settings* (Noordegraaf, 2015). In our example José works independently in a rural practice, while Laila is working within a larger urban practice. This might mean that Laila has to consider organizational rules about antibiotics prescribing that José has not. On top of that, the situational settings that public sector workers encounter are often complex. Laila and José might have to consider patients' beliefs about antibiotics, they have to examine vague symptoms that can

represent an array of medical conditions, and they must analyze treatments considering different medical backgrounds. In addition, the individual characteristics of public sector workers might matter for decision-making. Laila and José have different professional backgrounds, as well as personal characteristics. While José might be a seasoned GP, Laila has just started. While Laila is digitally savvy, José has trouble using new technology in his practice. This can have implications for how they behave.

These are topics that are highlighted in this dissertation. We consider the potential of behavioral sciences to support decision-making in different public sector contexts, for different public sector workers. In other words, we study how heuristics can be dealt with conceptually and empirically to benefit public sector decision-making across public sectors. In the next section, we will first describe how we define behavioral sciences. Second, we describe how behavioral public administration has come into being, and, third, we explore decision-making within public sector contexts. Fourth, we present the research questions and the research contributions. Finally, we end this introduction with an overview of the remaining chapters and how they relate to the research questions.

1.1 BEHAVIORAL SCIENCES BEYOND HEURISTICS

In 2008, ‘the Jedi approach to public policy’ (The Economist, 2008, para. 1), nudging appeared, seemingly out of nowhere, with the arrival of international bestseller ‘Nudge’ (Sunstein & Thaler, 2008). A nudge is “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler & Sunstein, 2008, p. 6). An example is changing the default of a choice; for instance, in the case of organ donation (Jachimowicz et al., 2019; Johnson & Goldstein, 2003). The default, which is what happens when a person does nothing, can be that a person donates their organs or not. Changing the default choice does not forbid any options or does not change economic incentives, but it can change the number of people who donate organs. Thus, people are gently ‘nudged’ into certain behaviors (Sunstein & Thaler, 2008). Empirical evidence showed promising results of novel techniques such as changing default options and using social norms to change behavior. Some of these interventions are shown in box 1.

Nudges are connected to libertarian paternalism, which suggests that people remain autonomous because they have the choice to divert from the suggested option (Sunstein & Thaler, 2003; Thaler & Sunstein, 2008). Yet, at the same time, people are steered to perform certain behaviors. Nudges have also been criticized for being manipulative and lacking transparency (Hansen & Jespersen, 2013). Nudges can be used for negative purposes as well, as has been emphasized by scholars on administrative burden, budge, and sludge (Moynihan et al., 2015; Oliver 2013b; Thaler 2018).

Box 1: Examples of successful nudges on citizens

Every year patients miss thousands of appointments. The consequences are that healthcare professionals do not use their time efficiently and that healthcare problems remain undetected. The non-attendance rate can be reduced by on average 34% by doing one simple thing: sending a reminding text message (Hasvold & Wootton, 2011).

Using three different nudges, people can be stimulated to save more. First, people can be asked to opt in to save a larger amount of money in the future. Second, saving rates can be linked to future pay rates and third, people can be enrolled until they are actively rolled out. This 'save more tomorrow' plan accounts for an increase in the average saving rates from 3.5% to 13.6% in 40 months (Thaler & Benartzi, 2004).

Not all people pay their taxes on time. Simplifying the letter in which the government communicates that tax still needs to be paid can be helpful. John and Blume (2018) found that this increased tax compliance by 4%.

Nudging builds on the use of heuristics and biases to change behavior. To understand what heuristics and biases are, we need to delve into behavioral economics. A core finding in behavioral economics (BE) is that people do not always act rationally, and often think and act through systematic automatic processes (Kahneman, 2003; Tversky & Kahneman, 1981). Behavioral economics centers around the idea that humans think through two interconnected, yet distinct, processes. This is referred to as 'dual process theory' (Chaiken & Trope, 1999; Evans, 2003; Evans & Stanovich, 2013; Kahneman, 2011). Throughout the dissertation, we will use the terms system 1 and system 2 as introduced by Stanovich and West (2000), to describe these two processes.

System 1 is the system that originates from before we evolved to be human. It can be found both in humans and animals. System 1 is impulsive, fast, automatic, and non-deliberative (Gawronski & Creighton, 2013; Kahneman, 2011). System 1 allows you to act without thinking when you are in danger. Thus, it is essential for survival. System 1 learns by associative learning processes. System 2 is younger and is viewed to be exclusively human (Evans, 2003). We use system 2 when we want to reflect. This system is somewhat rational and implies slow, deliberate, decision-making. It also allows for abstract thinking, something which system 1 cannot do.

In system 1, people use heuristics to make decisions. Heuristics are mental shortcuts that reduce complex tasks to simpler ones (Kahneman et al., 1982; Tversky & Kahneman, 1974). Shortcuts can be helpful or harmful. Heuristics can help, for instance, when there is no time to consider all options (Gigerenzer, 1991; Gigerenzer et al., 2008; Gigerenzer & Goldstein, 1996). However, these shortcuts can also lead to suboptimal decision-

making. In those cases, these heuristic are called biases (Kahneman et al., 1982; Tversky & Kahneman, 1974). A cognitive bias refers to when 'human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality' (Haselton et al., 2015, p. 968). Researchers have been particularly successful at identifying biases, with more than 175 biases detected so far (Benson, 2016). An example is status quo bias, which is a tendency to prefer the current situation over a possibly more positive future one (Nicholson-Crotty et al., 2019).

Nudging, as mentioned above, uses insights about heuristics and biases to influence behavior. Yet, scholars and practitioners can use insights about heuristics in ways other than merely nudging. For example, we can also try to counter the effect of biases by educating decision-makers about biases and by debiasing. Examples of educating people about biases can be found in healthcare, where several biases, such as framing biases, contribute to misdiagnoses, and informing healthcare professionals about this might prevent these biases from having an effect (Ludolph & Schulz, 2018; Smith & Slack, 2015). Through education, people are made more knowledgeable about their ways of thinking. This approach is advocated in works building on 'Nudge', such as Boost and Nudge-plus (Hertwig & Grüne-Yanoff, 2017; John et al., 2009). Related to this, we see that, instead of using biases to steer behavior, debiasing is attempted (Soll et al., 2015). Debiasing refers to mitigating the effect of the bias. Educating decision-makers partly overlaps with debiasing, as information provision can be used as a debiasing technique. Debiasing also refers to countering the effect of a bias by using other methods such as nudges and accountability (Aleksovska et al., 2019; Soll et al., 2015).

Apart from that, multidisciplinary fields, such as behavioral economics and behavioral public administration, increasingly have been drawing from behavioral theories beyond those which rely on insights about heuristics and biases (Ewert, 2020; Feitsma & Whitehead, 2019; Lodge & Wegrich, 2016; Sousa Lourenço et al., 2016). In fact, using the term 'nudge' often represents a broader field of behavioral sciences (Sousa Lourenço et al., 2016). We illustrate this by looking into the definition of 'behavioral economics'. No single definition of behavioral economics exists (Heukelom, 2012), but the concept can be understood from a narrow and a broad perspective (Ewert, 2020; Gofen et al., 2021). The narrow definition of behavioral economics specifies studying 'the cognitive processes that people employ when making decisions often systematically and, therefore, seemingly deliberately, violate the set of assumptions and axioms that underlie the dominant neoclassical model' (Oliver, 2013, p. 689). This definition specifically refers to those deviations from the classical rational economic model by using biased decision-making. A broad definition of behavioral economics combines psychology and economics to study human decision-making. From a broad perspective, behavioral economics can be viewed as one-half economics and one-half psychology (Heukelom, 2012). By extension, we can then define the 'behavioral sciences' element in behavioral economics as sciences that consider the effects of cognitive processes on decision-making. This concerns insights mainly from psychology, but also from for instance neuroscience (Adhikari, 2016; Kwon & Silva, 2019). Some examples of the use of behavioral sciences beyond using nudge are described in Box 2.

The body of literature on cognitive processes is extensive (see, for instance, Davis et al., 2014). Relating to public sector workers, in healthcare research the tradition of understanding behavior through psychology is particularly developed (John, 2018). Recently, these insights have been bundled into overarching frameworks (Birken et al., 2017). A well-known and commonly used framework is the Theoretical Domains Framework (TDF) (Michie et al., 2014). The TDF aim to integrate a wide plethora of behavioral theories. By consulting psychologists and implementation researchers, the TDF identified 128 constructs from 33 theories that were integrated into 14 dimensions (Cane et al., 2012; Michie et al., 2005). These dimensions can be 'emotions' but also 'beliefs about capabilities'. The Theoretical Domains Framework is presented in Appendix F. Similar insights are present in the Ability Motivation Opportunity (AMO) model, which is often used in HRM research (Appelbaum et al., 2000; Marin-Garcia & Tomas, 2016).

Box 2: Behavioral sciences beyond nudge

For doctors, the availability by which a diagnosis comes to mind can be misunderstood as evidence for how accurate the diagnosis is. This can happen when doctors have just encountered a similar case. Through a structured reanalysis of the case, this effect can be countered (Mamede et al., 2010) ~~Setting and Participants Experimental study conducted in 2009 at the Erasmus Medical Centre, Rotterdam, with 18 first year and 18 second year internal medicine residents. Participants first evaluated diagnoses of 6 clinical cases (phase 1).~~

Feitsma & Whitehead (2019) describe the use of behavioral sciences during a crisis in water pollution in Flint, Michigan. They describe that just focussing on using nudges to get people to not drink the polluted water was not sufficient. Rather the governmental behavioral change experts found that it was essential to tackle the mistrust in the government.

Before undergoing an operation, a patient can be tested on how biological functions are performing at that moment. Although these tests can be useful, they can also cost extra time and money, and can produce harmful effects. Using the a comprehensive behavioral model as an interview guideline, scholars found that antecedents for testing biological functions were beliefs about the usefulness of these tests (Patey et al., 2012) ~~the TDF is used to explore anaesthetists' and surgeons' perceptions of ordering routine tests for healthy patients undergoing low risk surgery. Methods: Sixteen clinicians (eleven anesthesiologists and five surgeons.~~

Apart from that, specific behavioral theories can be found in public administration research. These theories are based on psychology, but focus on one topic. An example is that a public sector worker can be motivated or not to implement a policy, depending on if the person agrees with the goal of a policy (Tummers, 2011). As such, the use of behavioral sciences in public administration is not new. Below, we explain more in-depth how behavioral sciences have been influencing public administration for many years and how the nudging movement has reshaped the role of behavioral sciences in public administration.

1.2 THE EVOLUTION OF BEHAVIORAL PUBLIC ADMINISTRATION

Herbert Simon stated that he wrote *Administrative Behavior* in the 1940s because:

"decision making is the heart of administration, and that the vocabulary of administrative theory must be derived from the logic and psychology of human choice"

(Simon, 1997, p. xi)

Nevertheless, integrating psychology and public administration was largely neglected until the late 2000s (Grimmelikhuijsen et al., 2017). Around that time, an increasing number of scholars started explicitly linking theories and methods from the field of psychology and economics to understand key public administration challenges. This development was conceptualized and described in public administration research by the behavioral public administration (BPA) stream (Grimmelikhuijsen et al., 2017), which has been growing ever since. As with behavioral economics, we can define behavioral public administration in a broad and a narrow way (Ewert, 2020; Gofen et al., 2021). A narrow view of behavioral public administration refers to the analysis of public administration from the micro-level perspective of individual behavior and attitudes, by using insights on *heuristics and biases* as far as the behavior of individuals and groups is concerned. A broad view, in line with Grimmelikhuijsen et al. (2017, p.45), would draw on insights from *cognitive processes* on the behavior of individuals and groups.

Studies relating to both the narrow and broad definition of behavioral public administration (BPA) have been present before the origin of BPA, even though they were not recognized as sharing a common denominator. Regarding the narrow view, and thus heuristics and biases, some examples are that researchers have focused on, for example, how group-processes can create bias. Groupthink was identified as having a negative effect on decision-making ('t Hart, 1998; Irving, 1971). This happens, because, rather than seeking the best argument, people are seeking concurrence. This is caused by social pressures and doubts that arise when other people voice their opinions in group settings (Hallsworth et al., 2018). Framing bias has also been identified as an important determinant of policy preferences (Tversky & Kahneman 1981). Stone (1997) explains that one can frame a

policy idea in different types of stories. Policy ideas can, for instance, be presented as a solution to a long-standing problem, or as preventing a future issue. Policy ideas can also be translated into different metaphors. Recent research empirically shows that this framing has effects. If crime is, for instance, referred to as a ‘beast’ or a ‘virus’ changes proposed solutions to social problems (Thibodeau & Boroditsky, 2015). Another example is the focus on in-group biases which refers to people favoring people in their own group over people outside of their groups (Aronson, 2018). Scholars studying representative bureaucracy have shown that the social and demographic traits of bureaucrats can influence the service they provide (Bishu & Kennedy, 2020). For example, street-level bureaucrats might favor people who belong to their own social and demographic groups.

We have also seen studies that use psychological insights that go beyond heuristics and biases. Lipsky (2010) was aware that street-level bureaucrats had to cope with pressures of the job. Because of their high workload, they have to make choices regarding which services they deliver to whom. These choices are sometimes guided by own attitudes towards target groups. For instance, if clients are perceived as less deserving, street-level bureaucrats might ration the supply of services (Baviskar & Winter, 2017). In addition, scholars have focused on how to steer employees’ decisions. This relates closely to human resources studies in the public sector (Knies et al., 2017). Public service motivation scholars pointed out that motivation to serve the public cause has been important for decisions (Perry & Vandenabeele, 2015). Thus, public service motivation connects more broad theories of motivation to the goals in the public sector (Vandenabeele, 2007). On top of that, red tape scholars have been directing attention at the perceived usefulness of rules and the responses to them (Bozeman, 1993).

We can conclude that behavioral sciences were already quite prominent in public administration before the explicit integration of psychology and public administration described by behavioral public administration. Why then did the behavioral public administration stream develop not that long ago? First, the behavioral public administration stream recognized what happened within practice and academia. In practice, policy makers increasingly started to use behavioral interventions, like nudges, to steer citizens behaviors (Thaler & Sunstein, 2008). These nudges came as an addition to already known behavioral change techniques: the carrot, the whip and the sermon, which represent respectively: incentives (carrot), mandates and bans (whip), and information campaigns (sermon) (Tummers, 2019a). The popularity of nudge can be explained by policymakers’ needs (Bogliacino et al., 2016). Policy makers seemed to be attracted to the idea of policy initiatives based on evidence that can be obtained through experiments (Straßheim et al., 2015). At the same time, the idea of cheap interventions was appealing to public organizations depleted of resources. Nudges are also connected to libertarian paternalism, which promises choice to heterogenous societies, in which one-size-fits-all solutions are not desirable (Lodge & Wegrich, 2016). This led to new organizations sprouting in public administration practice. We have for instance seen the rise of different Behavioral Insights (BIT) Teams in numerous countries including the United Kingdom, the Netherlands, the United States, Qatar, Lebanon and Singapore (Feitsma, 2019).

In academia, psychological insights have been increasingly used in public administration research. These insights were novel because they focused specifically on heuristics or nudges and used experimental methodology, as is often seen in psychological studies (Grimmelikhuijsen et al., 2017). This momentum resulted from increased attention for nudging in practice, but also from a long tradition of economics and psychology scholars figuring out how to explain human decision-making (John, 2018). In the background, this had been going on for years. Tversky and Kahneman had already pointed out cognitive biases as early as the 1970s (Tversky & Kahneman 1981; Tversky & Kahneman 1973).

Because the development of behavioral public administration is linked to the attention for using heuristics, it is not surprising that most studies in BPA focus on the influence of nudges and biases (Battaglio et al., 2018; Ewert, 2020). Yet, the field is now maturing by stressing different topics using psychological theories (Carrigan et al., 2020). Christensen et al. (2020) have, for instance, emphasized the influence of administrative burdens on citizens when general cognitive resources are low. Hattke et al. (2020) researched the emotional effects on citizens of dealing with red tape by using physiological measurements of emotions. Apart from that, most scholars have thus far focused on understanding and influencing citizens (Battaglio et al., 2018; Ewert, 2020). However, lately scholars are moving towards studying the behavior of public sector workers (Hallsworth et al., 2018). The latter comes with unique challenges, as these behaviors occur within public organizations. This implies that the next steps are to explicitly include public sector contexts in behavioral public administration studies. We explore this in the next section.

1.3 DECISION-MAKING IN PUBLIC SECTOR CONTEXTS

Most studies in behavioral public administration have been about how to understand and influence citizens decision-making (Battaglio et al., 2018). One common objection to nudging citizens is that public officials are themselves influenced by the same automatic processes (Feitsma, 2018; Lodge & Wegrich, 2016). This can also be an argument to redirect attention from citizens to people working in societal organizations (Sanders et al., 2018). Indeed, leading scholars have identified studies on the use of behavioral sciences within organizations as a promising new frontier (Chapman et al., 2021; Soman & Yeung, 2021). Thus far, empirical evidence has shown that public professionals use cognitive biases in delivering public service (Battaglio et al., 2018; Bellé et al., 2018; Butler & Brockman, 2011; Englich et al., 2006; Meier et al., 2015; Viscusi & Gayer, 2015).

These insights are confirmed by the Behavioral Insights Team (BIT) in the UK, who have published a report on how behavioral insights can be used to improve the performance of public sector workers (Hallsworth et al., 2018). In their report, 'Behavioral Government', they explain that biases influence at least three different processes essential to policy making: noticing, deliberating and executing. For noticing, for instance, framing bias is relevant for how public sector workers interpret information. An example is that policy support depends on whether a policy outcome is framed in terms of losses (e.g. the number of people who will die) or gains (e.g. the number of people that survive) (Bellé et al., 2018;

Sheffer et al., 2018). For deliberation, group processes can reinforce biases. Researchers found that in group discussions it matters quite a lot who speaks first (Ottaviani & Sørensen, 2001). Initial ideas get more attention than later ones, which might mean that important topics are overlooked. For executing, Hallsworth et al. (2018) identify that optimism bias can be a problem. This means that people overestimate their abilities, plans and future success. Liu et al. (2017) researched overconfidence among bureaucrats concerning climate change. They found that the more experience people have, the more likely they are to be overconfident. This can be problematic as overconfidence leads to taking more risks.

We, thus, see that behavioral sciences can be applied within the public sector. Yet, we do not know how specific public sector contexts matter for this application. Jones (2017) points out that, even though behavioral sciences often focus on individual phenomena, behavior rarely occurs within a vacuum. It rather occurs within different types of *contexts*: within group contexts, within organizations, and within the (inter)national overall architecture of the state (Roberts, 2020). The importance of including context has been highlighted in public administration (Pollitt, 2013). In fact, context has been identified as a missing link in explaining variation in results. The application of context to behavioral sciences can be very important as behavioral interventions are not always effective. Evidence on why this happens remains limited (Feitsma & Whitehead, 2019; Hummel & Maedche, 2019). This is recognized by the European Commission which emphasized that behaviorally informed policy interventions should be tested in different contexts to identify the effect of boundary conditions on the interventions (Sousa Lourenço et al., 2016).

Currently, different scholars focus on different levels of analysis. With the introduction of BPA as a micro-oriented stream, researchers warn about a great 'schism' approaching, with different streams focusing on different levels of analysis, working with theories and methods that cannot be combined (Moynihan, 2018). Scholars, therefore, call for bridging work to include not only the individual level, but also other levels of analysis (Ewert, 2020; Ewert et al., 2021; Roberts, 2020). Answering this call might be essential to connect research and scholars across public administration, and to harbor the potential of behavioral sciences to support decision-making processes in public sector context (Bhanot & Linos, 2020).

Context can encompass a multitude of aspects and is hard to define (Pollitt, 2013). It ranges from the rules of supranational decision-making bodies to group processes within a staff meeting. In this dissertation, we focus on elements from both institutional and situational settings that influence behaviors (Noordegraaf, 2015). *Institutional settings* can be understood by focusing on institutions, which are the formal and informal rules, as well as routines that shape decision-making (March & Olsen, 1983). These rules can refer to a 'logic of appropriateness', which comprise ideas about what behavior is appropriate, when and where (March & Olsen, 1983, 2019), and a 'logic of consequence', which is about preferences, measures, cause and effect. Institutional settings can be unique to the public sector as the public sector comes with specific accountabilities, organizational structures, and goals (Noordegraaf, 2015). Yet, these institutions can also differ by sector, organization, team and group.

CHAPTER 1

The interaction between behavioral interventions and institutions has been emphasized recently. Sunstein (2017) explains that the effectiveness of a default depends on the preferences that people have. He uses the example of the default to retain your own last name for men and women in the United States. While few men change their last name, most women do. This can be explained by the institutional context of the country and this might work differently in other countries. The interaction between institutions and behavioral factors has recently also been explored in the public sector. George et al. (2020) for instance explain that negative performance information matters more than positive information, but only when benchmarked with government standards. This implies that the presence of institutions in the form of rules interacts with the presence of negativity bias. Another example is when organizations facilitate group processes, which means that an individual's decision-making is no longer isolated. Group processes can also accommodate biases (Hallsworth et al., 2018). An example is where groupthink can make people focus on agreement with each other, rather than on reaching the best solution ('t Hart, 1998). As such, we expect institutions related to public organizations to influence decision-making.

Apart from institutional settings, the *situational settings* that public sector workers encounter are important for decision-making (Noordegraaf, 2015). Situations often differ in *complexity* (Noordegraaf and Abma 2003; Noordegraaf 2015). Complexity consists out of two dimensions: a technical and a normative dimension. Technical complexity refers to issues consisting out of many variables (Holland, 2014), that might be hard to measure (Noordegraaf & Abma, 2003). Imagine the construction of a single street versus the construction of a whole neighborhood. Normative complexity refers to issues being contested. This means that people do not agree upon what needs to be done, and information can mean different things for different people. For instance, building a refugee center in a residential neighborhood might be more normatively complex than the construction of a round-about in the same neighborhood. We know that people oppose to being nudged when they do not agree with the outcome (Loibl et al., 2018; Lucke, 2013; Tannenbaum et al., 2017). An example of the influence of normative complexity in the public sector is how previous opinions matter for the interpretation of information. Baekgaard et al. (2019) found that politicians interpret evidence based on their prior beliefs. They found that most politicians interpret performance information about schools correctly when schools were labelled as school A and school B. However, when they changed labels to "public" and "private" schools, interpretations change drastically. Public and private schooling is a contested topic in Denmark. Politicians then have trouble interpreting information correctly when the evidence is at odds with their own beliefs.

Within these different institutional and situational contexts, different public sector workers make decisions. These public sector workers can be surgeons, police officers, administrative workers, nurses, HR-professionals, or managers. These occupational identities seem to have an influence on decision-making. Scholars have recently found that debiasing politicians is harder than debiasing non-politicians, as politicians rely more on previous political attitudes and less on policy information (Christensen & Moynihan,

2020). In other words, the type of occupation seems to matter for how behavioral factors such as biases affect decision-making. Differences between public sector workers extend beyond occupations. Public sector workers are individuals that differ in their skills, their motivations and their beliefs. We know from earlier research that these factors shape decision-making as well (Michie et al., 2014).

Studying institutional and situational settings, as well as individual differences, is relevant as behavioral interventions, such as nudges, often follow a one-size-fits-all approach, in which they are directed at a single impersonal outcome. Including these insights might imply that we need to *adapt behavioral interventions to context and recipients*. For example, we can combine nudges with algorithms to create 'contextualized nudges' (Yeung, 2017). We can also adapt nudges to specific target groups. Sunstein (2012) has already identified personalized nudges as one of the future avenues of interest. This is supported by empirical research (Mills, 2020). For instance, using personalized nudges instead of general ones can increase nudging effectiveness when it comes to influencing internet users to choose stronger passwords (Peer et al., 2020). Personalizing behavioral interventions can also possibly be done by including algorithms (Mills, 2020).

1.4 RESEARCH AIM AND QUESTIONS

We have explained that the attention for heuristics and biases, and related to that nudging, has generated new research that explicitly integrates behavioral sciences and public administration. This is described by the behavioral public administration stream (Grimmelikhuijsen et al., 2017). The next step in this behavioral public administration stream is to advance the handling of heuristics conceptually and empirically. This can be done by providing conceptual clarity through explicitly disentangling narrow and broad views of behavioral sciences (Ewert, 2020; Gofen et al., 2021). This allows us to relate to different traditions of thought, to broaden our scope of research, and to connect the behavioral public administration stream with earlier behavioral research in public administration. We also need to include context in our handling of heuristics, as scholars have now turned to studying decision-making processes *within the public sector* (Battaglio et al., 2018; Hallsworth et al., 2018). This new stream of research requires incorporating the unique institutional and situational settings that different public sector workers encounter in empirical work. Thus, the research aim of this dissertation is to explore how behavioral sciences can contribute to the understanding, and subsequently, improve decision-making processes within the public sector.

The above leads to the following research question:

How can behavioral sciences facilitate the understanding and improvement of decision-making processes by public sector workers in public sector contexts?

To study our sub-questions, we apply different research methodologies. Behavioral public administration has been characterized by the use of experiments (Bhanot & Linos, 2020).

CHAPTER 1

In this dissertation, the value of experiments for the detection of causality is recognized (Gerber & Green, 2012). Yet, we use different methodology beyond the traditionally applied survey experiment (Battaglio et al., 2018; Bouwman & Grimmelikhuijsen, 2016). This dissertation uses a systematic scoping review to synthesize existing experimental evidence (The Joanna Briggs Institute, 2015). We also conduct a replication and extension study to generalize earlier findings as well as presenting a solution to them (Walker et al., 2017). In addition, we use mixed methods by adding qualitative data to experimental data. Mixed methods are rare in public administration. Nevertheless, they offer valuable insights that uncover empirical patterns, as well as their real-world mechanisms (Gilad, 2019; Groeneveld et al., 2015). Apart from using experiments, we conduct a qualitative study using semi-structured interviews. Qualitative studies are underused in behavioral public administration but can help study, for instance, novel topics (Gilad, 2019).

The main research question is divided into four sub-questions. Our four sub-questions represent different papers. Thus, this is a paper-based dissertation. For our sub-questions, we chose to focus on a range of topics that reflect the variety of our research interest. These four sub-questions reflect different angles of approaching the main question. As such, this dissertation uses a maximum variety approach for researching our main question (Flyvbjerg, 2006; Suri, 2011) and to explore the potential of behavioral sciences, while highlighting different applications of it to various practices, public sector workers and issues. This approach allows us to detect common themes exist that are present across different cases and that are meaningful through being present despite this heterogeneity (Suri, 2011).

We have chosen to vary on different types of behavioral theories, practices, public sector workers and issues. Behavioral theories on the one hand cover heuristics and biases (nudging, biases and debiasing). On the other hand, they extend to more general psychological knowledge in the form of general comprehensive models that aggregate psychological theories, but also specific psychological theories that focus on an individual topic. We also research different practices (service provision, public management, political deliberations), different types of public sector workers (public professionals, public employees, public managers, and politicians), and cover a range of different issues (evidence-based medicine, goal setting and performance appraisal, algorithmic public management and active participation in deliberations). An overview is shown in Table 1.

Table 1
Overview of sub-questions using a maximum variation approach

Sub-question	Behavioral theory	Practice	Public sector worker	Issue
1	Nudging	Service provision	Public professionals	Evidence based medicine
2	Biases and debiasing	Public Management	Public managers and employees	Goal setting and performance appraisal
3	General comprehensive models (Theoretical Domains Framework)	Political deliberation	Politicians	Active participation in deliberations
4	Specific psychological theory (Procedural justice)	Public management	Public employees	Algorithmic public management

Below, we present our sub-questions and explain what they study, as well as how they are related to each other and our main research question.

1. How can nudges be used to support service provision, more specifically the provision of healthcare by healthcare professionals?

Our first chapter and article focuses on how we can use insights on heuristics and biases to support decision-making in public service contexts. We first do this by studying how nudges can be used for service provision. We focused on healthcare as the only known nudges in service provision were applied there (Bourdeaux et al., 2014; Kullgren et al., 2018; Meeker et al., 2014). In fact, healthcare is a frontrunner in applying behavioral insights in the public sector, as is shown in multiple viewpoint articles (King et al., 2013; Mafi & Parchman, 2018; Vaughn & Linder, 2018). Specifically, we focus on the issue of evidence-based medicine, which is the use of current best evidence in making decisions about the care of individual patients by healthcare professionals (Sackett et al., 1996). However, due to information overload using the best evidence is challenging for health care professionals and so nudges might be a solution to this problem.

2. To what extent are public managers and employees affected by anchoring bias and how can this be reduced?

Our second chapter researches heuristics and biases from a different perspective than chapter one by studying how the effect of biases can be understood and reduced. We focus on how anchoring bias affects decision-making about those issues. Anchoring bias has been previously researched in other contexts and is one of the most robust biases out there (Furnham & Boo, 2011). We also test a consider-the-opposite intervention to reduce the effect of anchoring bias in management decisions (Mussweiler et al., 2000). Normally identifying biases is more popular than replicating their results in different context or proposing a solution (Battaglio et al., 2018; Larrick, 2004). We focus on researching the biases, especially of public managers and employees in public management. Also, we focus on the issues goal setting and performance appraisal; two core public management practices (Favero et al., 2016; Pedersen & Stritch, 2018a).

3. How can behaviors during political deliberations be understood, with an emphasis on active participation in online deliberations (during COVID-19)?

Our third chapter goes beyond heuristics and biases and uses a comprehensive model of psychological factors to make a behavioral analysis of how politicians actively participate in online deliberations during COVID-19. The comprehensive model we use is the Theoretical Domains Framework (TDF) (Cane et al., 2012; Michie et al., 2005). The TDF is based on an aggregation of 33 psychological theories. We use this model to study the issue of active participation by politicians in online deliberations as a result of the COVID-19 pandemic when many public decision-making bodies suddenly had to move deliberation into online spaces. We emphasize how individual differences in active participation in these deliberations can be explained and intervened upon. As such, we show how we can use comprehensive models of psychological factors to detect and solve problems related to novel issues in public sector decision-making.

4. How can the procedural justice of management decisions be increased, especially when algorithms are used for decision-making?

Our last chapter also concerns behavioral sciences beyond heuristics and biases and studies a specific psychological theory: procedural justice. Procedural justice is the extent to which people perceive the process of a decision as fair (Colquitt, 2001; Lind & Tyler, 1988). Procedural justice is a psychological reaction to the violation of norms (Lind & Tyler, 1988). We use procedural justice to study how public employees view the involvement of algorithms in managerial decisions in the public sector. This issue is important as algorithms are increasingly used for public decision-making (Veale & Brass, 2019). This chapter also gives us information about the future of behavioral public administration. Algorithms can contextualize and personalize nudges and act as a possible solution for the lack of heterogeneity in behavioral interventions based on individuals and contexts (Mills, 2020). They have also been identified by Kahneman and Sunstein as a possible solution for problems of ‘noise’ in decision-making (Kahneman et al., 2021). Therefore, we researched the potential of using algorithms for public management decisions by studying the effect of including algorithms in managerial decision-making on public employees’ perceived procedural justice.

1.4.1 ACADEMIC RELEVANCE

This dissertation studies how we can use behavioral sciences to support decision-making in the public sector. It contributes to a long tradition of research on decision-making processes in public policy process and service provision (Hill & Varone, 2021). We also join the behavioral public administration stream in explicitly connecting behavioral sciences and public administration. We go beyond current studies in this stream in three ways.

First, we emphasize a narrow *and* broad view of behavioral sciences, and by extension, of behavioral public administration. Currently most literature has focused on identifying

how heuristics and biases affect public sector workers. This is in line with the narrow view of behavioral sciences, which focusses on the influence of heuristics and biases (Ewert, 2020). Even though this has offered us information about what determines decision making (Battaglio et al., 2018; Dolan et al., 2012; Münscher et al., 2016), the current body of literature does not fully harness the potential of behavioral sciences. We contribute to this by emphasizing how we can use insights on cognitive processes beyond heuristics to study decision-making. More specifically, we do this by applying a comprehensive behavioral model to make a behavioral analysis politician's active participation in online deliberations (chapter 4) and considering how expectations about algorithms can shape procedural justice (chapter 5).

Second, we do not focus on influencing citizens, but rather join recent efforts concentrating on decision-making inside public organizations (Bellé et al., 2018; Hallsworth et al., 2018; Soman & Yeung, 2021). In doing so, this dissertation aims to connect individual foundations of behavior with context (Jilke et al., 2019; Moynihan, 2018; Roberts, 2020). This connects to recent reflections of the BPA stream (Bhanot & Linos, 2020; Moynihan, 2018), while also joining older debates in public administration about the inclusion of context (Gilad, 2019; Pollitt, 2013). We focus on institutional and situational settings. In chapter 2, we for instance focus on how institutional and situational settings influence a nudge's success. In chapter 3, we replicate an earlier study to see if the institutional settings matter for the effects of biases. In chapter 5, we include situational settings in our experimental design, by testing how complexity influences procedural justice and, in chapter 4, we conduct a behavioral analysis including institutional and situational settings in a political local government context.

Apart from institutional and situational settings, we also differentiate between individuals. In behavioral economics, differences between individuals have been identified as essential for decision-making processes, including reactions to behavioral interventions (Appelt et al., 2011; Sunstein, 2012). This extends to public administration where researchers have started to address the cognitive resources on the citizen side (Christensen et al., 2020; WRR, 2017), and the differences between politicians and public employees (Christensen & Moynihan, 2020; Vis, 2019). We contribute by studying differences in effectiveness across occupations and professionals (chapter 2 and 3) and show that individual characteristics such as skills and beliefs about consequences are important determinants when explaining decision-making (chapter 4).

Third, methodologically, we diverge from the earlier focus of behavioral public administration on using mostly survey experiments to test if decision-making in the public sector is biased (Battaglio et al., 2018). We do this by adopting different methods that can have unique advantages to understanding decision-making through behavioral science (Bhanot & Linos, 2020). In chapter 3, we replicate and extend an earlier study by conducting survey experiments to test if results are generalizable to other institutional contexts, and subsequently intervene upon this issue. This answers calls to strengthen empirical evidence on generalizability across different contexts (Walker et al., 2017). In

chapter 4, we demonstrate how experiments can be combined with qualitative data by using open questions. Mixed methods approaches in public administration are rare, but allows us to uncover what mechanisms determine how context has an effect on the relationship between algorithm-manager interactions and procedural justice (Gilad, 2019; Groeneveld et al., 2015).

Furthermore, we use methodology that is not experimental. In chapter 2 we show how a systematic scoping review can be used to identify nudges currently in use. These types of studies are necessary to provide accessible overviews of available evidence that can be easily used for future studies and to understand the state of the field (The Joanna Briggs Institute, 2015). In addition, we study a case through semi-structured interviews and identify the behavioral factors important to active participation in online deliberations by politicians. Qualitative methods are rarely explicitly combined with behavioral theories to study decision-making of public sector workers, with some notable exceptions (Feitsma, 2019; Jones et al., 2013). They are nevertheless particularly useful to help explore novel topics (Gilad, 2019).

By adopting this multi-method approach, we attempt to emphasize that although survey experiments have benefits, other types of methodology also fit within the behavioral public administration stream (Tummers, 2020). Where possible, we made our data open access to assure transparency and provide background data for those interested in more in-depth knowledge or replications.

1.4.2 SOCIETAL RELEVANCE

Our relevance for society is threefold. First, we offer practical tools to practitioners for applying behavioral sciences and change decision-making within the public sector. This follows up on the promise of the behavioral stream to provide simple interventions that can change human behavior for the better. Currently, the application of behavioral science has been mainly directed at citizens, while the public sector has been overlooked (Hallsworth et al., 2018). Application there is vital because public sector workers are susceptible to the same heuristics and biases as citizens (Feitsma, 2018; Lodge & Wegrich, 2016). Choice architecture is always present around public sector workers' decision-making, even without intervening, and can have adverse effects (Vaughn & Linder, 2018).

We shed light on applications within the public sector in various ways. We map existing nudges that are used on healthcare professionals for evidence-based medicine. By this we make an inventory of nudges in public organizations. This inventory can be used by practitioners wanting to recognize or apply nudging within their own organizations. Next, we develop a low-cost and low-in-intensity debiasing intervention to counter anchoring bias. This goes beyond merely identifying biases that influence public sector workers, which is often done in public administration research (Battaglio et al., 2018). We focus on countering the effect of bias by empirically testing an intervention. If successful, this can be applied in real world instances where anchoring bias is at play. We also show how

behavioral analyses can be used to understand decision-making. Behavioral analyses are rarely reported in public administration research but they can identify influential (psychological) factors on behavior. In addition, these factors can be connected to behavioral interventions (Michie et al., 2014).

Second, we indicate that contextual factors are important when applying behavioral sciences within the public sector. Behavioral sciences, and especially 'nudge' have been seen as an excellent solution to decision-making problems (Dolan et al., 2012; Thaler & Sunstein, 2008). Yet, from the replication crisis we learned that translating psychological insights to different contexts is not always easy and effective (Open Science Collaboration, 2015). The relevance of including context has been recognized by the European Commission, because it might increase the effectiveness of interventions (Sousa Lourenço et al., 2016). In this dissertation, we focus on how institutional, the situational settings, as well as types of public sector workers matter for the use of behavioral sciences in the public sector (Noordgraaf 2015).

Third, we focus on issues that hold their own specific relevance for society. We study evidence-based medicine, goal setting and performance appraisal, algorithmic public management, and active participation in online deliberations. Evidence-based medicine refers to healthcare professionals making decisions about patients using the best available evidence (Greenhalgh et al., 2014), which can lead to more cost-effectiveness and better health care (Masic et al., 2008). Evidence-based medicine has been regarded as an utopia because of information overload. The volume of information is so large that it is impossible for healthcare professionals to keep up (Allen & Harkins, 2005; Williamson et al., 1989). Medical decision-making is like any other decision-making subject to heuristics and biases (Croskerry et al., 2013a). Therefore, nudges could provide solutions (Croskerry et al., 2013b). This dissertation maps studies using nudges to provide an overview of their applications, unresearched gaps and usability in health-care. We also use this to draw inspiration for other service provision.

Next, we study the effect of biases on management practices in the public sector. Public management is essential for the functioning of the internal public apparatus, and adequate service provision (Favero et al., 2016; Knies et al., 2017). We address how anchoring bias influences goal setting regarding the responsiveness to citizens and performance assessments of public employees (Bellé et al., 2017, 2018). Goal setting and performance feedback relate to two types of core internal management practices in the public sector that contribute to public sector performance (Favero et al., 2016; Pedersen & Stritch, 2018a). We also study how we can debias anchoring effects. As such, we identify problems and also attempt to solve them.

In addition, we focus on the practice of political deliberation. Politicians represent interests from society. We contribute to knowledge about how we can safeguard representative democracy within online deliberative democracy, specifically through videoconferencing involving politicians (Albrecht, 2006; Brown, 2018; Friess & Eilders, 2015). We concentrate

on the issue of the inclusion of different politicians in deliberations by studying active participation of politicians in online videoconferencing deliberations. The COVID-19 crisis meant public decision-making bodies had to move to online spaces. The digital divide literature emphasizes how inequalities between individuals can exist as a result of these spaces (Scheerder et al., 2017; Van Deursen & Van Dijk, 2019). We, therefore, researched personal and contextual factors that influence active participation within online deliberation. These factors can subsequently be leads for behavioral interventions (Michie et al., 2014).

Finally, we study the potential of algorithms to take over or be involved in managerial tasks relating to human resource management within the public sector. Algorithms are increasingly used for managerial decision-making (Wesche & Sonderegger, 2019). This includes nudges that are based on algorithms (Yeung, 2017). This happens in the public sector as well. An example is predictive policing where algorithms identify hot spots of potential criminal acts (Meijer & Wessels, 2019). We, however, do not know how the inclusion of algorithms in managerial decision-making is viewed by public employees for different types of tasks. Our study contributes to understanding when algorithms are perceived as procedurally just by employees for management decisions, and subsequently when we should implement them. This study also sheds light on the potential of algorithms in general to contextualize and personalize behavioral interventions as well as solving problems of 'noise' (Kahneman et al., 2021; Mills, 2020).

1.5 OUTLINE OF THE BOOK

The introduction presents the research questions and sub-questions that are the foundation of the empirical chapters. Chapter 2 presents a systematic review mapping nudges aimed at strengthening evidence-based medicine by health care professionals. In chapter 3, a replication and extension study shows how anchoring bias is relevant for management decisions and assesses if we can debias management decisions. Chapter 4 considers online deliberation, and analyses which factors influence politicians' actively participation in online deliberations. Chapter 5 is about how the involvement of algorithms influences perceived procedural justice of public employees when it comes to management decisions. Table 2 presents an overview of the empirical chapters, their methodology and the sub- questions that the article covers.

Table 2*An overview of the chapters of this dissertation, their methodology and their status*

Chapter	Methodology	Sub-question	Status
2: Nudging healthcare professionals towards evidence-based medicine: A systematic scoping review	Systematic scoping review of 100 articles	1	Published in <i>Journal of Behavioral Public Administration</i>
3: Designing to Debias: Measuring and Reducing Public Managers' Anchoring Bias	Survey experiment: Replication and extension with 1221 public managers and employees	2	Published in <i>Public Administration Review</i>
4: Don't speak: exploring how behavioral factors affect politicians' participation in online deliberations	Semi-structured interviews with 8 politicians and 7 council clerks	3	Under review
5: The impact of using algorithms for managerial decisions on public employees' procedural justice	Two survey experiments combined with qualitative data with 235 public sector workers	4	Published in <i>Government Information Quarterly</i>

In chapter 6, the conclusion, we discuss these findings and answer the sub-questions, as well as the main research question. We also consider the academic and practical implications of this dissertation and we end with a conclusion.



CHAPTER 2

Nudging healthcare professionals towards evidence-based medicine: A systematic scoping review

This chapter is based on the following published article: Nagtegaal, R., Tummers, L., Noordegraaf, M., & Bekkers, V. (2019). Nudging healthcare professionals towards evidence-based medicine: a systematic scoping review. *Journal of Behavioral Public Administration*, 2(2). The co-author statement can be found in Appendix J.

ABSTRACT

Translating medical evidence into practice is difficult. Key challenges in applying evidence-based medicine are information overload and that evidence needs to be used in context by healthcare professionals. Nudging (i.e. softly steering) healthcare professionals towards utilizing evidence-based medicine may be a feasible possibility. This systematic scoping review is the first overview of nudging healthcare professionals in relation to evidence-based medicine. We have investigated a) the distribution of studies on nudging healthcare professionals, b) the nudges tested and behaviors targeted, c) the methodological quality of studies and d) whether the success of nudges is related to context. In terms of distribution, we found a large but scattered field: 100 articles in over 60 different journals, including various types of nudges targeting different behaviors such as hand hygiene or prescribing drugs. Some nudges – especially reminders to deal with information over-load – are often applied, while others - such as providing social reference points – are seldom used. The methodological quality is moderate. Success appears to vary in terms of three contextual characteristics: the task, organizational, and occupational contexts. Based on this review, we propose future research directions, particularly related to methods (preregistered research designs to reduce publication bias), nudges (using less-often applied nudges on less studied outcomes), and context (moving beyond one-size-fits-all approaches).

2.1 INTRODUCTION

Translating knowledge into clinical practice remains notoriously difficult (Grimshaw et al., 2012). For example, guidelines take on average more than 17 years to be adopted, and only about half of the guidelines ever achieve widespread clinical use (Bauer et al., 2015). Over the past 20 years, increased attention has been given to reducing the gap between evidence-based practice and policy. This has been described using various terms of which evidence-based medicine (EBM) is commonly used (Grimshaw et al., 2012). EBM refers to the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients (Sackett et al., 1996).

The limited effect of evidence on behavior might be caused by two challenges in evidence-based medicine. First, the success of evidence-based medicine has led to an overload of evidence being made available (Greenhalgh et al., 2014). Already in 1989, two out of three US physicians stated that the current volume of scientific information was too large (Williamson et al., 1989). This information overload makes it impossible for healthcare professionals to review the best available evidence for each individual case. In particular, the number of clinical guidelines is overwhelming. For example, a 24-hour audit in an acute care hospital identified 3,679 pages of national guidelines that were relevant to the immediate care of 18 patients (Allen & Harkins, 2005).

Second, general evidence has to be applied to individual cases by healthcare professionals (Greenhalgh et al., 2014; Junghans, 2007). Evidence-based medicine has been criticized for its emphasis on evidence as opposed to professional autonomy (Greenhalgh et al., 2014). A key stance in modern EBM is that knowledge should inform healthcare decision-making, but not necessarily dictate it. This is because, for example, in a very complex medical situation, a general guideline may do more harm than good. One can minimize harm by developing robust evidence-based guidelines that are sensitive to the complexity of patient care, but evidence should be used in combination with expert knowledge and patient needs (Dreyfus & Dreyfus, 2005; Sackett et al., 1996). As such, interventions to promote EBM should not be too restricting and retain the professional's autonomy to deviate.

Nudges might be a possible solution to the two evidence-based medicine challenges described above. A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sunstein 2008, p.6). Most nudges work through automatic cognitive processes and by changing the choice architecture of a decision. An example of a nudge is changing a default choice, for instance by changing the choice to donate organs from being "opt in" to being "opt out" (Johnson & Goldstein, 2003). Nudging can ease situations of information overload by making information easier to process, for instance by presenting guidelines in 'plain English' (Michie & Lester, 2005). Moreover, nudges have been claimed to leave room for professional autonomy as nudging does not remove freedom of choice (Sunstein & Thaler, 2003).

CHAPTER 2

Accordingly, it is not surprising that the potential of nudging healthcare professionals has been recognized (King et al., 2013; Mafi & Parchman, 2018; Vaughn & Linder, 2018). For instance, The Behavioral Insights Health Project at Harvard University was launched to improve medical decisions through tools and research from behavioral economics (Harvard Law School, 2018). Despite this, authors of recent nudge experiments claim to be aware of only a few other experiments nudging healthcare professionals (Bourdeaux et al., 2014; Kullgren et al., 2018; Meeker et al., 2014). Therefore, in this article we conduct a systematic scoping review to give an overview of reported nudges that aim to strengthen EBM. Systematic scoping reviews are used to map what evidence has been produced as opposed to systematic reviews that seek the best available evidence to answer a particular question (The Joanna Briggs Institute, 2015). As such, systematic scoping reviews are especially suitable when researching relatively unexplored fields dealing with broad concepts (Peters et al., 2015). They allow researchers to ask broad questions but adopt a systematic approach in mapping the literature. In our study, we answer the following questions:

1. What is the distribution (journals, countries, year of publication, usage of nudge terminology) of studies on nudges aimed at strengthening use of evidence-based medicine by healthcare professionals?
2. What nudges, aimed at strengthening the use of evidence-based medicine by healthcare professionals, are being applied towards which out-comes?
3. What is the design and methodological quality of experiments testing nudges aimed at strengthening evidence-based medicine by healthcare professionals?
4. To what extent is a nudge's success in strengthening evidence-based medicine by healthcare professionals related to the task, organizational, and occupational contexts?

The first research question concerns the distribution of studies on nudging healthcare professionals towards EBM. We aim to show whether studies are clustered in certain countries or journals, in which years the studies have been published, if studies use nudge-related terminology, as well as the usage of nudge terminology over time.

Answering the second research question will identify which types of nudges are already frequently used and which seem to be overlooked. This provides an overview of currently available studies on nudges aiming to strengthen evidence-based medicine by healthcare professionals by bundling the available evidence. In addition, this overview can be used as an inventory to design experiments to test new nudges.

It is important to note that our review does not aim to provide an exhaustive overview of nudge studies on healthcare professionals to date. Instead, our goal is to clarify the current state of nudges related to evidence-based medicine on healthcare professionals. As such, we focus on studies using nudge related terminology as well as studies referring to healthcare professionals' behaviors specifically to promote EBM or usage of evidence/guidelines.

Our third research question concerns the quality of, and any indications of, bias in the published studies. To assess this, we use a quality assessment tool for multiple study designs (ICROMS) (Zingg et al., 2016). This assessment of quality can inform future research designs and contribute to the methodological advancement of experiments in public administration (Bouwman & Grimmelikhuijsen, 2016; Margetts, 2011; Moynihan et al., 2017).

Our fourth research question concerns the relationship between the context and a nudge's success. We use a statistically significant difference in behavior in favor of the nudge intention as a proxy for success. Although nudging has been claimed to be highly effective (Szasz et al., 2017), some suggest that success might depend on the context (Gould & Lawes, 2016; Halpern et al., 2007; Liao et al., 2016; Mafi & Parchman, 2018). This relates to a key challenge of EBM: leaving sufficient professional autonomy to allow deviation depending on the applicability of the evidence in a specific context. We thus focus on "success" related to task, organizational, and occupational contexts in hopes of providing a first step toward informing theoretical models and practical decisions on the role of context in nudging.

The contribution of this study is to provide an overview of the current scope and methodological quality of studies on nudging medical professionals, with the aim of going beyond a one-size-fits-all approach to nudging by directing attention to the inter-play between nudges and context (Hallsworth et al., 2018; Jones, 2017). This links to a well-known criticism of the behavioral stream in public administration research and its study of micro-phenomena: that it has moved away from macro-phenomena and big questions (Moynihan, 2018). We follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach (Liberati et al., 2009) and use the PRISMA Extension for Scoping Reviews checklist (see Appendix B) (Tricco et al., 2018).

2.2 THEORY

Nudging has its origins in behavioral economics. A core foundation of behavioral economics is that humans mainly think through two overarching, but interconnected, processes. This is referred to as dual process theory (Chaiken & Trope, 1999; Evans, 2003; Evans & Stanovich, 2013). Here we will use the terms system 1 and system 2 as introduced by Stanovich & West (2000) to describe these two processes. Dual process theory has been supported by empirical evidence for separate brain structures (Rangel et al., 2008).

System 1 is described as a universal form of cognition present in both humans and animals (Evans, 2003). As such this system is the oldest of the two. Associative learning processes form processes in system 1. System 1 is generally automatic, fast and non-deliberative, allowing one to quickly make sense of a situation and identify how to act (Gawronski & Creighton, 2013; Kahneman, 2011). This system is essential in situations of critical survival. The other cognitive system, system 2, is much younger and is believed to be present only in humans (Evans, 2003). This system is somewhat rational and implies slow, reflective thinking and deliberate decision-making. System 2 permits abstract thinking that cannot be achieved by system 1.

System 1 is characterized by the use of heuristics. Heuristics essentially reduce the complex tasks in assessing probabilities and values to simpler tasks (Lewis, 2008; Tversky & Kahneman, 1974). These heuristics are often very helpful and may help healthcare professionals to avoid errors, for instance in medical decision making (Marewski & Gigerenzer, 2012). Heuristics, however, sometimes lead to systematic errors which are labelled biases (Benson, 2016; Tversky & Kahneman, 1974). Cognitive biases occur when 'human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality' (Haselton, Nettle, and Murray 2015, p. 968). An example is confirmation bias, which represents the seeking or interpreting of information that is in line with existing beliefs (Nickerson, 1998).

Here, we are not considering the cognitive processes, but rather the techniques designed to affect decision-making using processes from system 1. These techniques are often called nudges. For instance, a default might use the status quo bias to nudge people into staying in a savings plan (Thaler & Benartzi, 2004). A key characteristic of nudges is that they do not rule out any option nor change economic incentives, thereby safeguarding professional autonomy. We accompany our description with a nudge taxonomy. Different taxonomies exist which reflect different preferences in thinking about nudges (e.g. Dolan et al. 2012; Johnson et al. 2012; Susan Michie et al. 2011; Sunstein 2014). Münscher, Vetter, & Scheuerle (2016) developed a nudge taxonomy, with the goal of creating mutually exhaustive and exclusive sets, on the basis of 127 documented examples of empirically tested interventions. We have adopted this taxonomy because of its systematic approach.

Münscher et al.'s (2016) nudge taxonomy has three main categories: decision information, decision structure and decision assistance. Decision information refers to changing the way information is presented without changing the options themselves. This can, for instance, refer to presenting guidelines in plain English or providing a social reference point (Allcott, 2011; Michie & Lester, 2005). Decision structure is about altering the arrangement of options and the decision-making format. This amounts to changing how alternatives are presented. An example is reducing the number of options that can be easily selected, or changing the effort needed to make a certain decision by changing the default (Johnson & Goldstein, 2003). Decision assistance refers to closing the intention-behavior gap (Sheeran, 2002). Here, people are provided with tools aimed at helping them follow up their intentions. Examples are reminders and asking people to specify when and where they will complete an action (Hagger & Luszczynska, 2014).

2.3 METHODOLOGY

2.3.1 SCOPE OF REVIEW

For a study to be included in the review, it had to deal with nudges that were applied to healthcare professionals on the individual level to promote evidence-based medicine. We focused on encouraging decisions that are seen as appropriate, that is, in accordance with evidence (Proctor et al., 2011). Whether an intervention constituted a nudge was

determined using the taxonomy by Münscher, Vetter, & Scheuerle (2016). Studies that focused on adherence to practice guidelines were considered eligible since practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances” (Field and Lohr 1990, p.8). We chose to include guidelines because they offer instructions on different behaviors related to clinical practice such as which diagnostic or screening test to order, how to provide medical or surgical services and hand hygiene (Woolf et al., 1999). Although we have not registered this review, all the codes used are provided online in JBPA's Dataverse.

Only reports on experiments were eligible for inclusion. Experiments were seen as comparing the effects of two or more interventions (The Cochrane Collaboration, 2018) and included randomized controlled trials (RCT), before and after studies (BA) and interrupted time series (ITS). Only studies written in English were considered. We did not impose constraints on the year of publication.

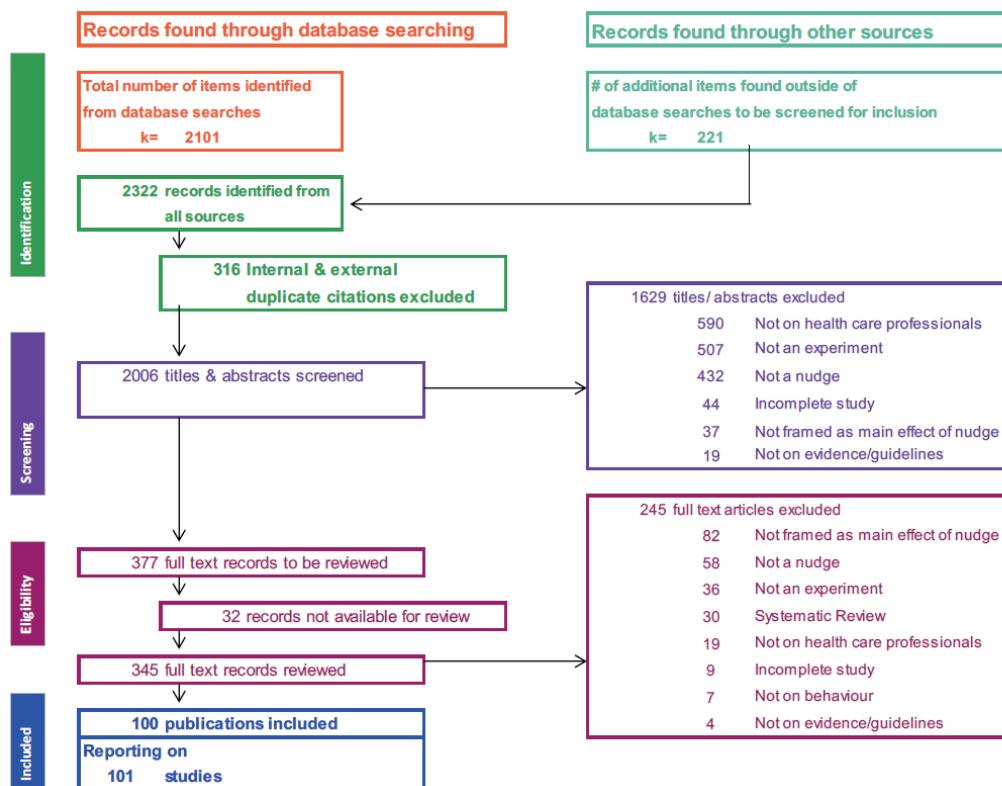
2.3.2 SEARCH STRATEGY AND STUDY SELECTION

To find eligible studies, we used four methods (Cooper, 2010). First, we searched the Ovid MED-LINE, PubMed, and PsycINFO databases using combinations of the term ‘nudging’ with ‘experiment’, ‘physicians’, ‘guidelines’, or similar terms (producing 65% of the total articles retrieved). The specific details of this search strategy are shown in Appendix A. Second, we searched for studies in several top journals that, according to our first search, publish articles concerning nudges on healthcare professionals, namely The Lancet, The British Medical Journal (BMJ), Annals of Internal Medicine, the Journal of the American Medical Association (JAMA), Implementation Science, and BMJ Quality and Safety (producing 25% of the articles retrieved). Third, we scanned relevant overview articles including those identified in the database searches (for example, Szaszi et al., 2017) to find further eligible studies (10% of total articles retrieved). Finally, we consulted experts to check the list of publications and identify any potentially overlooked studies (1% of total articles retrieved). The search process was concluded on May 25th, 2018.

CHAPTER 2

Figure 1

PRISMA Flow Diagram, Based on Workbooks for Systematic Reviews in Excel (VonVille, 2018)



The study selection process is shown in Figure 1. First, we screened 2,322 publications by scanning the abstracts and titles in a blind manner (i.e. concealing authors and journals). We checked if our inclusion criteria (such as topic and language) were met and checked for duplication. Of these 2,322 articles, 377 were deemed potentially eligible and we then read the full texts of these publications. During the full text readings, studies were either excluded or coded in full. The codes used were critically appraised on multiple occasions and refined accordingly. Tabulations and summaries are based on these codes. All the included publications are listed in the Supplement. After this final check, we were left with 100 publications, which report on 101 studies and 124 interventions.

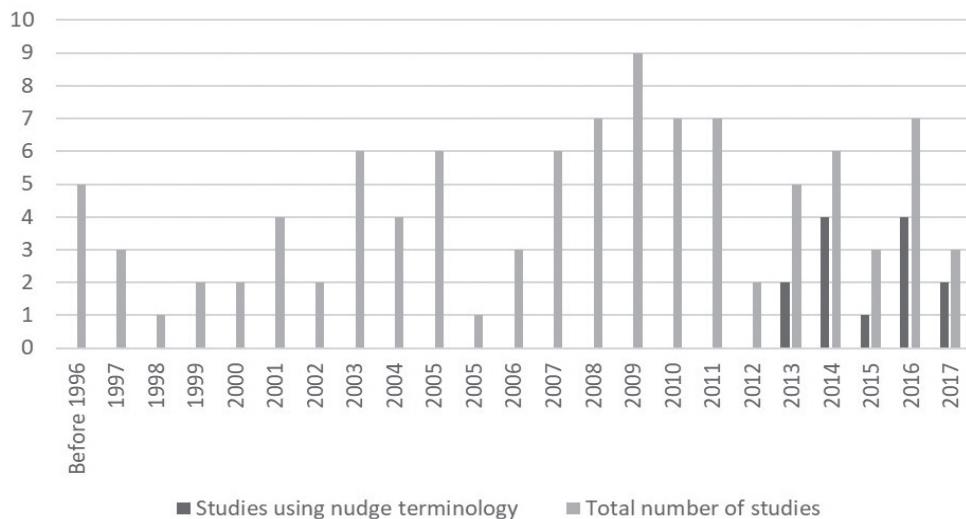
2.4 RESULTS

2.4.1. DISTRIBUTION OF THE STUDIES

We first discuss the distribution of the studies (RQ1). We found that most studies were conducted in Western countries, with a strong dominance of the United States (59% of

all studies) (e.g. Flanagan et al. 1999; Schwann et al. 2011; Tierney et al. 2005) and 10% in the United Kingdom (e.g. Bourdeaux et al. 2014; King et al. 2016; Weir CJ et al. 2013). Only a few studies were from non-Western countries, such as Kenya (Zurovac et al., 2011) and Taiwan (Hung et al., 2008). This suggests that a Western perspective dominates, which could have important implications as a country bias might be present. This might also influence the external validity of the findings, raising questions as to how applicable they might be in non-Western settings. Further, we found that all the included studies were conducted in a single country, indicating a lack of cross-country comparisons.

Figure 2
Years of publication and usage of nudge terminology



The articles included in the systematic scoping review were published in 64 different journals. Most were published in healthcare journals such as the Journal of the American Medical Informatics Association (8%) (e.g. Field et al. 2009; Rood et al. 2005; Sequist et al. 2005) and the Journal of the American Medical Association (8%) (e.g. Dexter et al. 2004; Feldstein et al. 2006; Junghans 2007). Besides these healthcare journals, articles were also found in more general behavioral science or implementation journals, such as in Implementation Science (4%) (e.g. Beidas et al. 2017; Kousgaard et al. 2013; Verbiest et al. 2014). In Figure 2, we show the publication years and indicate whether nudge-related terminology was used. We coded a study as containing nudge terminology if we found terms such as "nudge", "behavioral economics" or "choice architecture". Figure 2 indicates that there was a peak in publications around 2007 to 2011, but that nudge terminology was not used until 2013.

2.4.2 NUDGES AND TARGETED OUTCOMES

The studies included in our review used various nudges as shown in Table 3. Our search highlighted a diverse field with at least four published interventions in every category. Many studies (42%) concerned reminders and/or making information visible (e.g. Filippi et al. 2003; Förberg et al. 2016; Murtaugh et al. 2005). Studies in the largest category often used a form of computerized decision support that provides alerts, based on available guidelines, about the appropriate-ness of a certain decision. As Table 3 shows, the other categories were much less common. For instance, we found only five studies that facilitated commitment (Casper, 2008; Erasmus et al., 2010; Kullgren et al., 2018; Meeker et al., 2014; Verbiest et al., 2014). A detailed description of all the interventions by category can be found in the files for this article uploaded to the JBPA Dataverse.

We found that the largest category contained interventions aimed at changing prescribing habits (30%) (e.g. Flanagan et al. 1999; Larsen et al. 1989; Strom, Schinnar, and Bilker 2010). Other studies were on laboratory tests or diagnostic image ordering (26%) (e.g. Gill et al. 2009; Kahan, Waitman, and Vardy 2009; Kucher et al. 2005) or on hand hygiene (18%) (King et al., 2016; Kwok et al., 2016; Nevo et al., 2010). A few studies addressed other behaviors such as medical handovers (Messing, 2015) and providing cognitive behavioral therapy (Beidas et al., 2017).

The type of nudge being used seems to be related to the desired outcomes. Nudges on hand hygiene mostly involved changing option-related efforts (36%) (e.g. Chan, Homa, and Kirkland 2013; Nevo et al. 2010), such as by changing the location of hand hygiene dispensers. We did not find any studies on hand hygiene that involved nudges in the form of making information visible, providing reminders, or changing defaults. Studies on prescribing mostly involved making information visible or providing reminders (54%) (e.g. Buisling et al. 2008; Hicks et al. 2008; Martens et al. 2007).

Table 3*Applied nudge categories and techniques (based on Münscher et al., 2016)*

Nudge category	Number	Example
<u>A. Decision information</u>		
A1 Translate information	9 (7%)	Emphasizing consequences for patients of proper hand hygiene (Grant & Hofmann, 2011)
A2 Make information visible	23 (19%)	Suggesting alternatives when clinicians propose antibiotics (Meeker et al., 2016)
A3 Provide social reference point	7 (6%)	Showing general practitioners that they prescribe more antibiotics than their peers (Hallsworth et al., 2016)
<u>B. Decision structure</u>		
B1 Change choice defaults	9 (7%)	Changing the default for tests from optional to preselected (Olson et al., 2015)
B2 Change option-related efforts	8(6%)	Putting medical tools in line of sight (hand hygiene dispensers) (Nevo et al., 2010)
B3 Change range or composition of options	10 (8%)	Grouping tests on order forms or displaying them individually (Kahan et al., 2009)
B4 Change option consequences	4 (3%)	Asking for accountable justifications (Meeker et al., 2016)
<u>C. Decision assistance</u>		
C1 Provide reminders	28 (23%)	Putting reminders on operating room schedules (Patterson, 1998)
C2 Facilitate commitment	5 (4%)	Hanging poster-sized commitment letters including photographs and signatures (Meeker et al., 2014)
Other (Multifaceted)	21 (17%)	Providing cues through posters and stickers in a schematic breast shape with space for recording three mammography referrals on charts (Grady, Lemkau, Lee & Caddell, 1997)
Total (n)	124	(This is higher than the number of studies as some studies addressed multiple nudges.)

Changing prescribing habits was also nudged by providing social reference points (10%) (e.g. Denton et al. 2001; Hallsworth et al. 2016; Kiefe et al. 2001). Studies related to ordering habits mostly nudged by making information visible or providing reminders (51%) (e.g. Bindels et al. 2003; Lo, Matheny, and Seger 2009; Roukema et al. 2008) but also by changing the range or composition of options (18%) (e.g. Kahan, Waitman, and Vardy 2009; Poley et al. 2007). No studies on changing option-related efforts were found related to prescribing or ordering.

Nudges were administered in different types of environments. Most were applied in digital environments (66%), followed by nudges on paper (15%). Some nudges altered the position or presentation of objects in the physical environment (6%). The remaining nudges involved changing the environment, for instance by adding a clean smell (e.g. Birnbach et al. 2013), were delivered by people, or were delivered in multiple or unspecified ways. Most nudges (70%) that were applied in digital environments aimed at changing ordering or prescribing behaviors (e.g. Melnick et al. 2010).

2.4.3 QUALITY OF STUDIES

To answer the second research question, we assessed the methodological quality of the studies using ICROMS (Zingg et al., 2016): a single-step approach for assessing the quality of studies with multiple study designs. ICROMS provides criteria for assessing the quality of different study designs while allowing scores to be compared. Below, we show the scores for the different categories in Table 4. ICROMS scores for all the included studies are in the JBPA Dataverse files for this article. For those studies with randomized controlled trials (RCT), controlled before and after studies (CBA) and controlled interrupted time series (CITS) designs, the average score met the minimum required level. The mean scores in the non-controlled before and after studies (NCBA) and non-controlled interrupted time series (NCITS) categories were below the minimum required score, with none of the NCBA studies meeting the minimum threshold. This gives an indication of the lower quality of such non-controlled before and after studies (NCBAs). However, these numbers only tell part of the story about study quality. For instance, in many studies key information was often not given, making it hard to evaluate the risk of bias. We now zoom in on specific criteria where there is clearly room for improvement in the two categories with the most studies: RCT and NCBA studies. In RCTs, allocation concealment was generally rated poorly (57% of the maximum possible score on average). A solution for this would be to have the allocation carried out centrally by an independent third party (as in Van Wyk et al. 2008). Moreover, many studies could suffer from selective outcome reporting since, in many instances, no study protocol was provided and it was not explicitly stated whether studies were selectively reporting or not (on average, these studies scored 58% of the maximum possible score). The situation could be improved by authors opting to preregister experiments which would also address publication bias problems (Stern & Simes, 1997). The NCBA studies scored particularly poorly with only one study (Creedon, 2005) justifying the sample chosen or carrying out a baseline measurement to prevent selection bias. Here, researchers could pay more attention to how their sample might create a bias in the results, for instance by comparing sample demographics to the demographics of the population being studied. Furthermore, very few studies attempted to justify the lack of a control group (score of 15% of the maximum possible) and only one (O'Connor et al., 2009) attempted to mitigate the effects of not having a control group. This indicates that there is a risk of bias in most studies that use such a design.

Table 4
ICROMS scores per category

Design category	Number of studies	Mean score (range); max possible score	Minimum required score
Randomized controlled trial (RCT)	70 (68%)	22 (5 – 27); 32	22
Non-controlled before and after study (NCBA)	17 (17%)	16 (11 – 21); 30	22
Non-controlled interrupted time series (NCITS)	5 (5%)	18 (10 – 23); 30	22
Controlled before and after (CBA)	6 (6 %)	18 (10 – 24); 30	18
Controlled interrupted time series (CITS)	3 (3%)	18 (18 – 19); 30	18
Total	101 (100%)	N/A	N/A

2.4.4 SUCCESS OF NUDGES BY CONTEXT

Our third research question focused on the contextual conditions under which nudges are successful. The studies included in our review are highly heterogeneous. We, therefore, conducted a narrative synthesis. We used significant changes in behavior in the preferred direction as a proxy for success (Szasz et al., 2017). In addition to the type of nudges that are successful, we wanted to explore to what extent the context matters in the success of nudging. Nudges are potentially dependent on three types of context: the task, organizational and occupational contexts. Most studies (65%) reported positive results. The categories with the highest percentages of positive outcomes were changing option-related efforts (88% of studies reported success, for instance Chan, Homa, and Kirkland, 2013) providing social reference points (71%, for instance Hong et al. 1990) and using a combination of nudges (76%, for instance Hulgan et al. 2004). The categories with the highest percentages of mixed outcomes were facilitating commitment (40%, for instance Kullgren et al. 2018) and changing choice defaults (22%, for instance Ansher et al. 2014). Change option consequences had the highest percentage of null outcomes (50%, for instance Beidas et al. 2017), followed by translating information (44%, for instance Jousimaa et al. 2002). Very few negative effects were reported (a notable exception being Dexter et al. 2004), which could be due to publication bias. Further details on the interventions are provided JBPA Dataverse files for this article. In terms of context, the task at hand clearly matters. In the reviewed studies, nudging to promote hand hygiene was most successful (77%). A reason for this could be that the need for hand hygiene is widely accepted (Luangasanatip et al., 2015) and nudging might be less successful for other outcomes whose desirability is questioned. For instance, the effect of action planning on care to encourage smoking cessation was particularly apparent among GPs who had already intended to implement this activity but had not yet done so routinely (Verbiest et al., 2014). Mixed results were more commonly found for nudges related to ordering tests and diagnostic imaging (24% of studies reported mixed results). Sequist et al. (2005) provide an example of mixed findings in noting that the success of the intervention they studied depended on the service being recommended and the particular disease. This indicates that professionals will deviate from the nudging intention if they find the promoted action inappropriate.

Sometimes nudges are designed so that they adapt to reflect individual cases. These nudges are based on algorithms. For instance, in one study IF-THEN rules were created based on guidelines (Martens et al., 2007). These rules generate specific reminders for relevant cases, but not for others. This contextualization of the nudge can be beneficial in reducing problems created by applying general guidelines to individual cases. However, such applications are limited. Martens et al. (2007) further indicated that they were not certain whether complex recommendations always translated into meaningful reminders. In addition, some physicians rebelled at the notion of a computer telling them how to manage their patients (Tierney et al., 2003).

Nudges may well work differently in different organizational contexts. Our review showed that the most successful nudges were reported in hospitals (74% of studies in hospitals report positive results). A study by Kiefe et al. (2001) noted that physicians in rural settings were less likely to improve treatment by responding to feedback. This could be because rural physicians are more autonomous. Helder et al. (2012) indicated that not only the organization, but even the type of unit or shift can influence the results. They reported an overall positive effect for a screen-saver intervention, but no effect when calculated for the night shift alone. They suggest that nudges might work better in highly visible situations and not so well when people operate individually.

The effectiveness of nudges depends on the occupational context, meaning that success depends on the professional that is working with the nudge. For instance, academic physicians might be more aware of guidelines, influencing their reaction to nudges (Martens et al., 2007; Tannenbaum et al., 2015) and newly qualified residents might be more susceptible to nudges than more experienced physicians (Cummings et al., 1982; Fogarty et al., 2013). This is an indication that public professionals, depending on their level of professionalization, react differently to nudges.

2.5 DISCUSSION

As far as we are aware, this is the first systematic scoping review to map studies on nudging healthcare professionals towards applying evidence-based medicine. In this review, we have studied the distribution, the nudges and the targeted outcomes, the methodological quality, and the influence of context on the nudges' success. Based on our results, we draw four conclusions. We relate these conclusions to EBM challenges in dealing with information overload and applying professional autonomy when applying general guidelines to individual cases.

2.5.1 DISTRIBUTION OF STUDIES

Our first research question was about the distribution (journals, countries, year of publication and nudge terminology) of studies. We have three main conclusions. First, most studies are conducted in Western settings, and all of them in a single country. This raises questions about the external validity of the findings. Future studies could be conducted in other country settings. Second, we found studies in 64 different journals. This emphasizes the need for scoping reviews such as this one to bundle available evidence. Third, healthcare professionals 'have been nudged' since 1974. However, nudge-related terms were not used until 2013, indicating that interventions have only recently been recognized as nudges.

2.5.2 TYPES OF NUDGES AND TARGETED OUTCOMES STUDIED

Our second research question was about what types of nudges have been applied and towards which outcomes. We found that studies testing nudging are more widespread than

often claimed (Bourdeaux et al., 2014). Some nudges, such as reminders in computerized decision-support systems, are studied more often than many others, such as using defaults. The focus on reminders makes sense as reminders address the EBM challenge of coping with information overload: reminders make relevant information easily available to healthcare professionals at point-of-care. Nevertheless, other nudging forms can also mitigate information overload. Nudges could for instance make existing guidelines easier to use by simplifying their format (John & Blume, 2018; Michie & Lester, 2005).

Apart from information overload, nudges target 'irrational' behavior by healthcare professionals and use cognitive biases to change behavior. For instance, nudges can facilitate commitment to close the intention-behavior gap or change defaults in ordering systems (Ansher et al., 2014; Kullgren et al., 2018). These nudges might be especially useful when barriers other than information overload have been identified. For instance, for fairly general guidelines about hand hygiene, the location of hand hygiene dispensers has been described as a main barrier to compliance by nurses (Sadule-Rios & Aguilera, 2017). There are, however, only a few related studies and further research is needed. Furthermore, the nudges studied mainly focus on outcomes related to ordering, prescribing, and hand hygiene. Future research could test existing EBM nudges in less researched areas, such as administration and medical handover. In designing new studies, one should be aware that some nudges are more applicable to certain behaviors than others. For instance, it is not surprising that we did not find any studies using a default-type nudge to encourage hand hygiene since having clean hands by default is unachievable. In comparison, we also found few studies reminding healthcare professionals to wash their hands – a nudge that seems highly feasible. Further, even without actively nudging, the design of current systems might have an influence on performative behaviors. Choice architecture is always present and, if options are not displayed, this will influence the choices people make (Tannenbaum et al., 2015). Therefore, we would encourage critical reviews of existing choice architectures (Vaughn & Linder, 2018).

2.5.3 METHODOLOGICAL QUALITY OF STUDIES

Our third research question focusses on assessing methodological quality. The methodological analysis indicated that many studies were only of moderate study quality. Researchers could improve methodological quality to reduce the risk of bias and simultaneously increase the validity of the study outcomes. We would urge quality improvements by making small changes, such as ensuring allocation concealment is carried out by a third party, and also by making larger changes, such as by preregistering experiments. In terms of non-controlled before and after studies, more attention should be paid to the potential bias introduced by sample selection, and the omission of a control group should always be justified. Moreover, we often found studies were unclear as to what choices had been made, and why. Collectively, we should therefore strive to increase our reporting standards. We suggest using reporting guidelines and checklists, such as the Consolidated Standards of Reporting Trials (CONSORT) statement (Moher et al., 2010).

2.5.4 NUDGES IN DIFFERENT TASK, ORGANIZATIONAL AND OCCUPATIONAL CONTEXTS

In our fourth research question, we highlight the role of three contextual conditions for success: task, organization and occupation. We first note that 65% of published studies report success (i.e. statistically significant improvements). This could be due to publication bias, which is characterized by an aversion to publishing studies with null results (Ferguson & Heene, 2012). Here, we suggest preregistering experiments as a partial solution (Nosek & Lakens, 2014). Nevertheless, the 65% of 'successful' studies in this paper is considerably below the 83% successful intervention rate reported in a more general systematic scoping review of nudges (Szasz et al., 2017). We can offer two reasons for this. First, publication bias could be less widespread in studies dealing with evidence-based medicine than studies about nudges in general. Second, it could be that nudges are less successful in EBM due to other factors such as study design or contextual factors. We summarize the influence of task, organizational, and occupational contexts below.

First, we see that the targeted task is important in determining the success of a nudge. This could be because tasks that are widely accepted, such as hand hygiene, are more suitable to nudging. Related to this, some outcomes would seem less appropriate to nudging. In a clinical context, appropriateness de-pends to a large extent on outcomes. For example, Patel, Volpp, and Asch (2018) state that reducing the default duration of opioid prescriptions may make sense in acute conditions, as often seen in an emergency department, but may be inappropriate for clinicians caring for patients with chronic pain. This example further stresses the importance of carefully considering the behaviors being nudged.

Some nudges present contextualized information based on algorithms. This diminishes the problem of using general guidelines in individual cases, as nudges become customized to specific clinical scenarios. The question is, to what extent should nudges be contextualized for specific tasks? Evidence-based medicine has been criticized for overly focusing on algorithmic rules that oversimplify clinical realities (Greenhalgh et al., 2014). In line with this, complex clinical realities might not always be suitable for nudges, as nudges always involve some form of simplification, either through IF-THEN rules or by targeting a quite general outcome such as reduced antibiotic prescribing ('Thou should not prescribe antibiotics for cases of flu'). Here, we see that EBM nudging suffers from a similar problem to that of applying heuristics: simplifying complex realities can be beneficial, but not all situations can be easily simplified. We would therefore advise practitioners and authors to consider nudge-task fit and assess impressions of the complexity and appropriateness of the targeted behaviors with specialized healthcare professionals.

Second, the organizational context seems to have an influence. Physicians in a large city hospital have been found to react differently than a rural physician (Kiefe et al., 2001). Nurses during the night shift might not be influenced by nudges that are effective during the day shift (Helder et al., 2012). More research is needed on how working autonomously, in teams, and/or under various levels of visibility can make nudges more or less effective.

Third, the occupational context is important. Less experienced doctors are, for instance, more inclined to accept a default than experts (Fogarty et al., 2013; Martens et al., 2007). More information on the interplay between professionalism and nudges would be useful. In terms of algorithms, it has been shown that if people are experts, or believe they are experts, they tend to follow decision rules less often and as a result perform worse (Arkes et al., 1986). Does the same occur if 'experts' override nudges such as default options, or are these experts a necessary counterforce to the nudge? Overall, we see a need for future research to focus on the implications of task, organizational, and occupational contexts for nudges and thus to move away from a one-size-fits-all view of nudging. Instead, the focus should be on how the context of public professionals matters in nudging (Jones, 2017).

2.5.5. LIMITATIONS

The present review has several limitations. First, we cannot be certain that this review covers all nudges related to evidence-based medicine by healthcare professionals. In systematic scoping reviews, the trade-off between breadth and comprehensiveness is often reported as a challenge (Pham et al., 2014). Our search strategy focused on behavioral aspects in healthcare, seeking studies referring to nudge-related terms and studies referring to healthcare professionals' behaviors to promote EBM. In this sense, our study encompasses an already broad spectrum of studies that goes beyond those using nudge terminology but might nevertheless have overlooked studies using other terms (Szasz et al., 2017). Especially for reminders, there is already a large body of literature (for an overview see Cheung et al. 2012). Our findings could also be skewed due to publication bias. We attempted to address this by explicitly asking experts to add unpublished studies, but it is possible that some relevant studies have been overlooked.

Second, the heterogeneity of the studies meant that we could not conduct a meta-analysis. Instead, we have provided a systematic scoping review (Szasz et al., 2017). We recognize that even though heterogeneity is a strong argument against conducting meta-analyses, our systematic scoping review is limited because it does not consider effect size, sample and other relevant measures (Ioannidis et al., 2008). Moreover, in this study we use a statistically significant difference in behavior in the direction of the nudge intervention as a proxy for success. Future research could also carry out meta-analyses of specific categories in those areas where there is sufficient homogeneity in the published studies. For some nudging categories, such as reminders, meta-analyses of their effects on healthcare professionals already exist and provide more detailed information on their effectiveness (Cheung et al., 2012). Other nudging categories, such as using defaults, need additional studies with similar designs in order to assess their effectiveness with healthcare professionals.

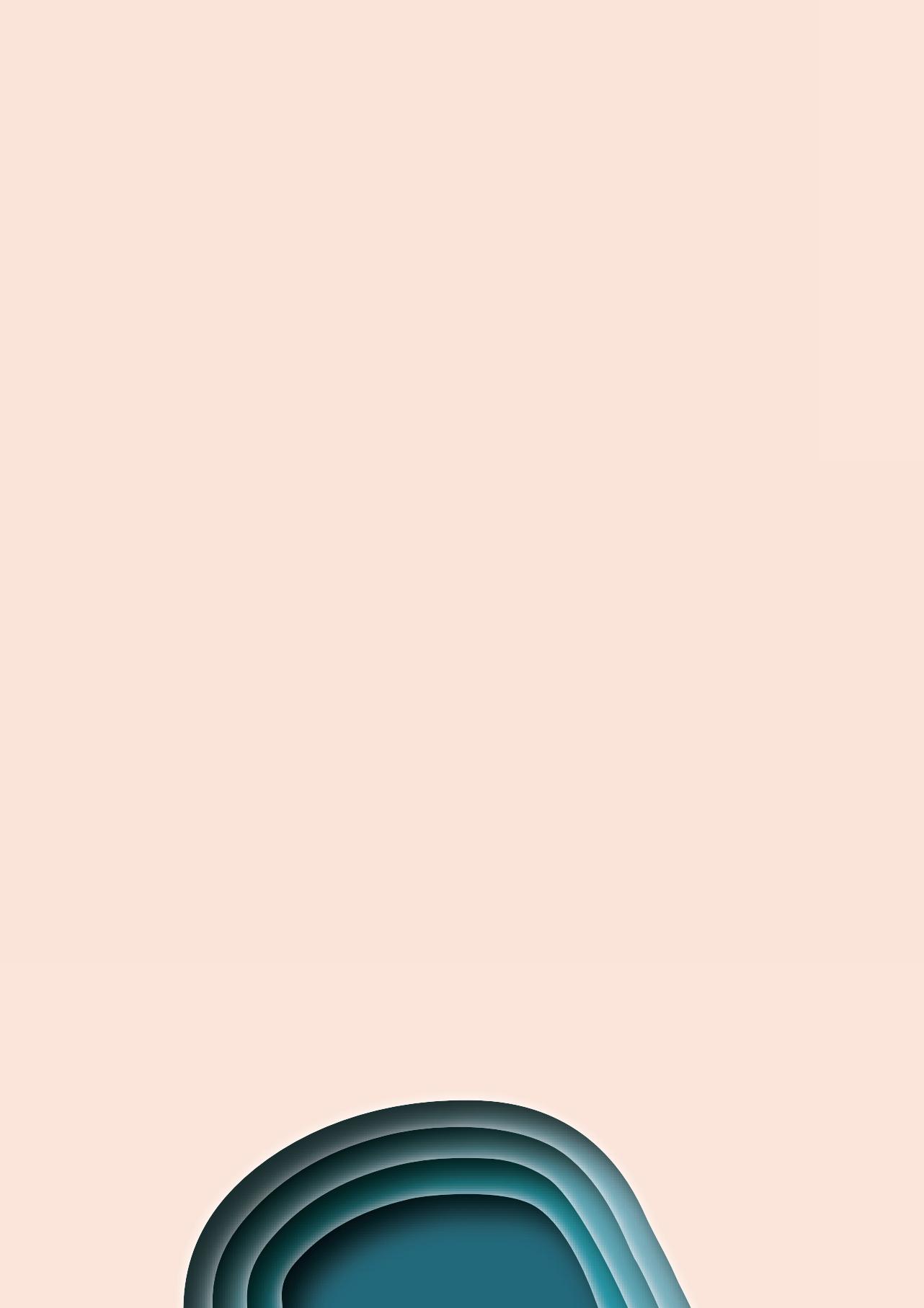
Third, 'success' can also be evaluated in terms of other outcomes. O'Connor et al. (2009) for instance stated that while most changes in order sets were beneficial, order set changes were also associated with an unintended overall increase in ordering night-time sedation.

Tierney et al. (2003) noted that physicians and pharmacists found the nudge intrusive and time consuming. Although such issues are beyond the scope of this review, these reports highlight the importance of not only studying significant differences, but also evaluating the impact on professionals' attitudes and unintended negative consequences.

Fourth, we categorized interventions in the choice architecture category we found most fitting. However, we found the choice architecture categories by Münscher, Vetter, & Scheuerle (2016) to be not entirely exclusive of each other. Therefore, we advise scholars looking for interventions in a particular category to review the related categories in the JBPA Dataverse files for this article as well. Despite these limitations, we do believe that we have shed new light on the scope of the nudging field and identified possible avenues for future research.

2.6 CONCLUSIONS

The aim of this research was to expose the current state of research on nudge interventions designed to promote evidence-based medicine by healthcare professionals. We found more than a hundred studies in over sixty journals and identified ten distinct nudging categories associated with outcomes ranging from hand hygiene to prescribing. Furthermore, we found that nudges have been used since the 1970s, despite nudge terminology not appearing until 2013. Reminders that deal with information overload are used the most often. However, further studies on less reported nudge categories that could also mitigate information overload, such as the effect of simplifying existing guidelines, are required. We also need more studies that explore outcomes beyond hand hygiene, image ordering and prescribing, as well as assessments of current choice architectures. Our methodological assessment identified considerable room for improvement in the identification of success, through better study design and more detailed reporting, with suggestions made related to allocation concealment and preregistration. Future research should also consider the roles of task, organizational, and occupational contexts in theoretical models regarding the design of nudges, thereby moving beyond one-size-fits-all approaches.





CHAPTER 3

Designing to debias: Measuring and reducing public managers' anchoring bias

This chapter is based on the following published article: Nagtegaal, R., Tummers, L., Noordegraaf, M., & Bekkers, V. (2020). Designing to debias: Measuring and reducing public managers' anchoring bias. *Public Administration Review*, 80(4), 565-576.
The co-author statement can be found in Appendix J.

ABSTRACT

Public managers' decisions are affected by cognitive biases. For instance, employees' previous year's performance ratings influence new ratings irrespective of actual performance. Nevertheless, experimental knowledge of public managers' cognitive biases is limited, and debiasing techniques have rarely been studied. Using a survey experiment on 1,221 public managers and employees in the United Kingdom, this research (1) replicates two experiments on anchoring to establish empirical generalization across institutional contexts and (2) tests a consider-the-opposite debiasing technique. The results indicate that anchoring bias replicates in a different institutional context, although effect sizes differ. Furthermore, a low-cost, low-intensity consider-the-opposite technique mitigates anchoring bias in this survey experiment. An exploratory sub-group analysis indicates that the effect of the intervention depends on context. The next step is to test this strategy in real-world settings.

3.1 INTRODUCTION

Appropriate practices by public managers are essential for public sector performance and, as a consequence, for well-functioning bureaucracies (Favero et al., 2016). However, decisions made by all human agents are subject to predictable cognitive biases (Tversky & Kahneman, 1974). Cognitive biases occur when "human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality" (Haselton, Nettle, and Murray 2015, p.968). For instance, people make different decisions when information is framed negatively than when it is framed positively (Bellé et al., 2018; Tversky & Kahneman, 1981).

Empirical evidence shows that public managers use cognitive biases in public sector decision-making (Battaglio et al., 2018). Availability bias and anchoring bias are important, for instance, in macroeconomic forecasts that provide policy information for managing the U.S. economy (Krause, 2006), framing matters for performance evaluations of organizations and individuals (Belardinelli et al., 2018), and local election officials' overconfidence in their own judgment affects their technology preferences (Moynihan and Lavertu 2012). Despite this recognition, the body of knowledge on the effects of cognitive biases on public managers' decision-making is still limited.

This article reports on a research strategy of initially replicating two experiments by Bellé, Cantarelli, and Belardinelli (2018, 2017) on cognitive biases in the public sector. These experiments represent two types of core internal management practices in the public sector: goal setting (e.g., establishing the maximum number of days within which to respond to emails) and performance feedback (e.g., performance ratings) (Favero et al., 2016; Pedersen & Stritch, 2018a).

The aim of this replication is threefold (Jilke et al., 2017). First, the replication extends the generalizability of earlier results. Experimental results of one population and context might not generalize to another (Lykken, 1968). That is why empirical generalization is essential to test the robustness of findings (Walker et al., 2017). Second, replications reduce the chance of false positives (Ioannidis, 2018). Third, the influence of context might be tested through replications. Therefore, experiments should be tested across politico-administrative contexts.

In our case, most evidence indicates that anchoring bias is very robust across contexts (Furnham & Boo, 2011). Nevertheless, new research has shown that the effect of biases might be dependent on institutional context (Christensen, 2018; Dudley & Xie, 2020; Holm, 2017). In other words, cognitive processes apply to individuals, but they happen within institutions (Jones, 2017). This institutional perspective on the effects of cognitive processes, including biases, remains understudied.

Therefore, we are interested in how the effects of cognitive biases can be generalized to other institutional contexts. In this case, this research compares the original results from

Italy with results from the United Kingdom, as these two countries represent different politico-administrative regimes (Pollitt & Bouckaert, 2011). Replications remain rare in research even though the importance of replications is evident (Brandt et al., 2014; Pedersen & Stritch, 2018b; Walker et al., 2017). Thus, this article contributes to building a body of knowledge, rather than relying on one experiment to substantiate claims.

Further, this research focuses on testing a low-intensity, low-cost debiasing technique. The impact of biases on decision-making has led scholars to suggest that biases should be taken into account when designing the architecture of jobs and tasks (Bellé et al., 2018; Vaughn & Linder, 2018). However, in general, research demonstrating biases is more widespread than research on solving bias-related problems (Bhanot & Linos, 2020). It seems to be “more newsworthy to show that something is broken than to show how to fix it” (Larrick 2004, p.334). An explanation for this is that cognitive biases are robust and debiasing is notoriously difficult (Lilienfeld et al., 2009). Nevertheless, given the effects of biases, debiasing has the potential to increase public sector performance, and therefore it should be on the agendas of practitioners as well as researchers. This article is focused on testing a debiasing technique that can be easily applied in practice.

This article concentrates on one specific cognitive bias: anchoring bias. Anchoring bias refers to the tendency to estimate unknown quantities by using an initial value (Tversky & Kahneman, 1974). This bias is chosen for two reasons. First, anchoring bias has real-world implications for public management. Examples are quantitative evaluations of employees’ performance, where last year’s employee ratings affect this year’s ratings (Bellé et al., 2017); decisions on academic promotions, where performance criteria inform decisions, irrespective of performance (Chen & Kemp, 2015); negotiations, where initial offers anchor negotiation outcomes (Guthrie & Orr, 2006); and evaluations, where political and historical performance results label current performance as either a failure or success (Holm, 2017).

Second, anchoring is notoriously robust, and previous attempts to debias anchoring in public management have not succeeded (Cantarelli et al., 2018; Furnham & Boo, 2011). However, there is a promising strategy that could be translated into public sector practices: consider-the-opposite. Although this strategy has been tested in other settings and has proven to be successful (Lord et al., 1984; Mussweiler et al., 2000), no reported experiments exist that test consider-the-opposite as a low-cost, low-intensity intervention to debias decisions in public management. This leads to the following research question: *Does anchoring bias affect public management decisions across institutional contexts, and can anchoring bias in decision-making be mitigated through a low-cost, low-intensity consider-the-opposite strategy?*

The article is structured as follows: First, it elaborates on the theoretical background by discussing cognitive biases and specifically anchoring bias. The article also expands on debiasing techniques and argues that a consider-the-opposite strategy is an appropriate strategy for debiasing anchoring effects for public management decisions.

Second, the method is explained: a survey experiment involving 1,221 public managers and employees, part of which is a replication. Third, the authors discuss the results and, fourth, their implications for public management practice and scholarship are considered.

3.1 THEORY

First, the article explains the most important concepts used in this study.

3

3.1.1 COGNITIVE BIASES

To understand cognitive biases, it is necessary to start with dual process theory. Dual process theory is a broad cognitive theory about the workings of the human mind and shows that people make decisions through two interconnected cognitive "systems" (Chaiken & Trope, 1999; Evans & Stanovich, 2013; Kahneman, 2011). System 1 allows people to make decisions rapidly, automatically, and intuitively. System 1 is the oldest of the two systems and not exclusive to humans. System 1 is particularly useful in dangerous situations as it allows us to act without consciously having to think. The other system, system 2, is slower and more reflective. Applying system 2, people can go beyond our first hunch and consider more complex factors that may be relevant to the problem at hand.

In system 1 decisions, shortcuts, or heuristics, are used. Heuristics simplify complex decisions (Tversky & Kahneman, 1974). In many situations, this is beneficial, but it can sometimes lead to systematic biases. Researchers have been particularly successful at identifying biases, with more than 175 biases detected so far (Benson, 2016). Notable examples are status quo bias, which refers to the tendency of people to stick to the current situation (Kahneman et al., 1991), and confirmation bias, which refers to people interpreting or looking for information that is concurrent with their existing beliefs (Nickerson, 1998). Cognitive biases influence decisions made by public managers and employees (Battaglio et al., 2018; Grimmelikhuijsen et al., 2017).

3.1.2 ANCHORING

This article focuses on anchoring bias. As noted earlier, anchoring bias has many real-world effects, including influencing performance ratings, and it is notoriously robust (Bellé et al., 2017; Furnham & Boo, 2011). The current dominant view is that anchoring works by activating anchor-consistent information (Furnham & Boo, 2011). As such, anchoring is an association-based bias (Larrick, 2004). Association-based biases make some information more available in the mind than other information during decision-making, creating selective accessibility (Strack & Mussweiler, 1997). In essence, anchoring bias thus induces a reference frame in a person, a certain set of thoughts, which makes it difficult to consider alternative possibilities (Koehler, 1991).

In this article, two earlier anchoring experiments are replicated (Bellé, Cantarelli, and Belardinelli 2017, 2018). These experiments represent two core categories of internal

management practices (Favero et al., 2016; Pedersen & Stritch, 2018a). Internally focused managerial practices aim to change employees' behavior. They fall into four broad categories: they set goals, build trust, increase participation in decision-making, or provide performance feedback.

The first experiment illustrates a goal-setting practice. Specifically, the scenario used here concerns establishing a maximum number of days that employees have to respond to citizens' emails (Bellé, Cantarelli, and Belardinelli, 2018). Public organizations rely on written digital communication to send information to citizens and stakeholders (Faulkner et al., 2018). Nevertheless, decisions about what constitutes a timely response could be influenced by anchors that might be irrelevant (Bellé, Cantarelli, and Belardinelli, 2018).

The second experiment focuses on providing feedback on public employees' performance through performance ratings (Bellé, Cantarelli, and Belardinelli, 2017). Performance ratings are a common form of performance appraisal in the public sector and are widely studied in public administration. However, performance ratings have been shown to be prone to errors and biases (Tummers, 2017). Anchoring bias has proven to be very robust across different contexts (Furnham and Boo, 2011), and there is no reason to believe otherwise in the case of internal management practices across institutional contexts.

This leads to the following hypothesis:

Hypothesis 1

Participants in the high-anchor replication groups will report estimates that are significantly higher than estimates from participants in low-anchor replication groups.

3.1.3 DEBIASING

As debiasing is fairly novel in public administration research, an overview of the debiasing literature is provided. There are two main overarching debiasing categories (Croskerry et al., 2013b; Keren, 1990; Larrick, 2004; Soll et al., 2015). One category includes strategies that *modify the environment*, which either makes the bias irrelevant or mitigates its effect. People could, for instance, use nudging or hold public employees accountable for their decisions (Aleksovska et al., 2019; Nagtegaal et al., 2019). The second category involves *modifying the decision maker*, which can be done by providing education on the bias at hand and the consequences that a bias might have and/or providing tools to mitigate the effect of the bias (Lilienfeld, Ammirati, and Landfield, 2009). Strategies that modify the decision maker are grounded in a two-model system of reasoning that is related to dual process theory. This model assumes that people first make an intuitive judgment about a situation with system 1 and that this judgment can be corrected by reflective and more effortful thinking with system 2 (Milkman et al., 2009; Morewedge et al., 2015).

This article focuses on *modifying the decision maker* and the cognitive processes used to make a decision. Interventions that are informative and use system 2 have been

argued to preserve individual dignity by allowing individual agency, and therefore they are preferred by the people affected by interventions (Sunstein, 2016). Not all strategies that modify the decision maker are equally promising. For example, Cantarelli, Bellé, and Belardinelli (2018) tested an educational debiasing intervention in public service. This intervention did not overcome anchoring bias.

Others claim that a combination of education and tools is needed to achieve effective debiasing (Adame, 2016; Morewedge et al., 2015; Wilson & Brekke, 1994). This can be very intensive in terms of resources and undesirable in the public sector, which is characterized by low resources (Lipsky, 2010). However, others claim that offering only corrective tools could be effective (Larrick, 2004). Here, one strategy seems especially promising to debias anchoring: the consider-the-opposite strategy. This strategy is discussed next.

3.1.4 CONSIDER-THE-OPPOSITE STRATEGIES

First, it is important to realize that no debiasing strategy works on all types of biases (Croskerry et al., 2013a). In practice, biases work through different mechanisms, and some biases might work through multiple mechanisms at once (Larrick, 2004). In this article, we attempt to mitigate anchoring bias by using a consider-the-opposite strategy because this strategy is a good fit for association-based biases such as anchoring. By its nature, anchoring creates a cognitive reference frame, making it hard to consider alternative thoughts (Koehler, 1991). Consider-the-opposite is a technique to break through this frame and open the door to alternative reasoning (Mussweiler et al., 2000).

The consider-the-opposite strategy has been found to be effective in dealing with biases, such as confirmation bias (Anderson, 1982; Hirt & Markman, 1995), framing (Cheng et al., 2014), and the anchoring effect (Adame, 2016; Lord et al., 1984; Mussweiler et al., 2000). The consider-the-opposite approach is administered mostly by simply asking people to list reasons why the anchor value is inappropriate (Adame, 2016; Kennedy, 1995; Lord et al., 1984; Mussweiler et al., 2000). In the past, consider-the-opposite has been tested, for example, on attitudes toward the death penalty, judging individuals' personality traits (Lord et al., 1984), probabilities of a correct diagnosis (Arkes et al., 1988), estimating the value of a car, and estimating the probability of election outcomes (Mussweiler et al., 2000). Our experiment tests the consider-the-opposite approach on two scenarios representing two core internal management practices in the public sector. Moreover, our experiment uses real public managers and employees as a sample, increasing the external validity.

A further change from previous studies is that this research tests an online low-cost, low-intensity, and thus scalable, version of the consider-the-opposite strategy. Previously, most consider-the-opposite experiments have involved the researcher being present (Lord et al., 1984; Mussweiler et al., 2000). However, recent research has shown that consider-the-opposite can work as an online intervention, provided that it is part of a

training process (Adame, 2016). The following hypotheses are formulated:

Hypothesis 2

Participants in the low-anchor consider-the-opposite group will report estimates that are higher than participants in the low-anchor replication group.

Hypothesis 3

Participants in the high-anchor consider-the-opposite group will report estimates that are lower than participants in the high-anchor replication group.

3.3 METHODOLOGY

Experiments frequently require a trade-off between control and internal validity and external validity and realism (Druckman et al., 2011). In this experiment, we opted for a controlled design with high internal validity. Therefore, we argue that if we do not find an effect here, we probably will not find an effect in more realistic scenarios. Our two scenarios are about establishing the maximum number of days within which an employee has to respond to citizens' inquiries (Bellé, Cantarelli, and Belardinelli, 2018) and about giving a performance rating to an employee (Bellé, Cantarelli, and Belardinelli, 2017). This research conducts replication experiments with the goal of achieving empirical generalization (Walker et al., 2017). Consequently, the research design, measures, and analysis in the original experiment are used.

The original experiments were administered in Italy. This experiment is conducted on public managers and employees in the United Kingdom. The United Kingdom and Italy are interesting cases because they represent different politico-administrative regimes with different histories, rules, and practices. For instance, these countries differ in terms of many politico-administrative variables, including state structure and administrative culture (Pollitt & Bouckaert, 2011). Italy is increasingly decentralized, whereas the United Kingdom has a more centralized structure. Apart from that, Italy has been described as a mild adopter of New Public Management (NPM) practices, which makes it a country in which managerial and traditional models are mixed (Nitzl et al., 2019). In contrast, the United Kingdom is a heavy adopter of NPM.

Specific to the scenarios, in the United Kingdom, the provision of information to citizens has been documented in the Freedom of Information Act (FOIA) of 2000 (Worthy, 2010). The FOIA dictates a maximum of 20 days for responding to citizens' requests for most governmental organizations (Information Commissioner's Office, 2019). In Italy, a FIOA was passed in 2016, establishing a maximum of 30 days to respond (Repubblica.it, 2016). The original studies were conducted in Italy just after adoption of the FIOA law in June–July 2016 (Bellé, Cantarelli, and Belardinelli 2017, 2018).

Concerning the performance feedback scenario, both the United Kingdom and Italy use performance assessments and feedback as a regular part of human resource practices (OECD, 2012b, 2012a). Both countries use performance criteria such as interpersonal skills, activities undertaken, and improvement of competencies. In both countries, performance assessments are of high importance for remuneration and career advancement. As such, performance feedback is relevant in Italy as well as the United Kingdom. Nevertheless, differences also exist. As stated earlier, Italy and the United Kingdom differ in their adoption of NPM practices. Research has shown that NPM can affect the ways in which performance information is used (Nitzl et al., 2019). This makes the United Kingdom an interesting case with which to assess empirical generalization.

The study was conducted using the crowdsourcing tool Prolific. Crowdsourcing refers to the use people participating in an online environment to complete a variety of tasks (Sheehan, 2018). The benefits of crowdsourcing are large-scale recruitment of participants in a short time, low costs, and access to a broader population. The downsides are a lack of control over the context in which the respondent takes the survey, loss of naivety, and possibly ethical problems because no set standards for payment exist (Palan & Schitter, 2018; Shank, 2016). Prolific, however, has been designed with the academic community in mind and therefore addresses these downsides, for instance, by not allowing researchers to pay less than an established minimum wage.

We used the pre-screening option on Prolific to select people from the United Kingdom who are public employees. To get paid, the participants had to complete the whole study. We used the "forced response" option in Qualtrics so that participants could not continue with the survey unless the questions were answered. This resulted in 1,221 respondents who were randomized and 1,202 respondents who finished the whole survey. The percentage of missing data is thus very small (1.5 percent). Cases with missing values for either the grouping variable or the dependent variable were excluded from the analysis of that dependent variable.

Replications need a highly powered sample to confirm that the effect of the original study is significant (Brandt et al., 2014). The sample size was chosen based on a pilot experiment of the whole study, involving 16 people. Based on this pilot, a power analysis was conducted, which led to an estimation of 282 respondents needed per group to corroborate a one-tailed hypothesis for our smallest effect (Cohen's $d = 0.21$). This was in line with the sample size of the original experiments.

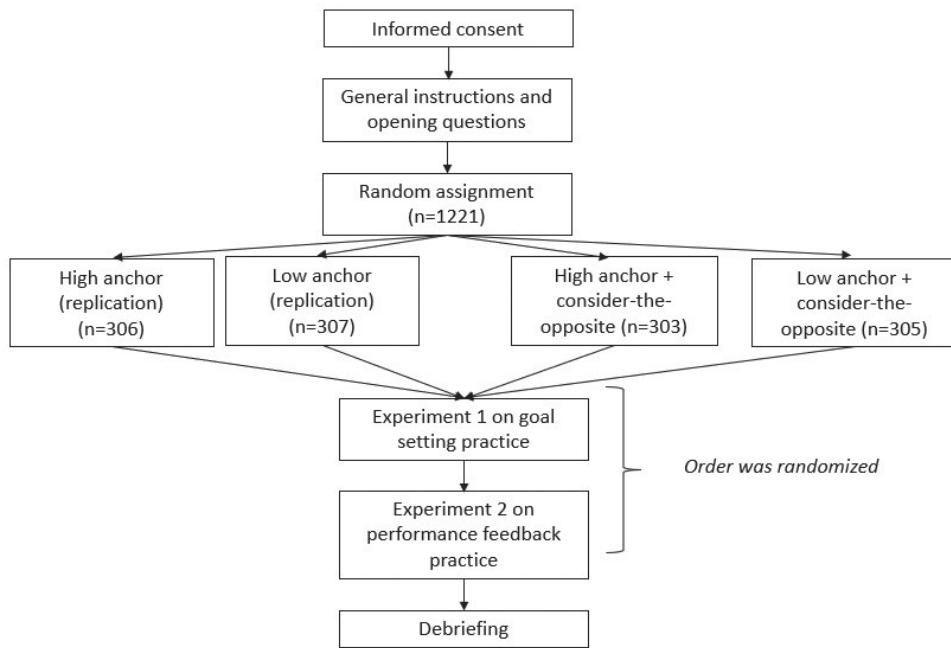
The debiasing intervention is based on earlier consider-the-opposite experiments (Adame, 2016; Lord, Lepper, and Preston, 1984; Mussweiler, Strack, and Pfeiffer, 2000). Respondents were asked to state two reasons why the anchor is inappropriate. The direction in which the anchor was inappropriate was specified. In other words, if the anchor was too low, people were asked to explain why the anchor was too low. If the anchor was too high, people were asked to state why it was too high. Therefore, the intervention was simple, low-cost, and low-intensity, and it could be applied even when the researcher was not present. All scenarios and interventions are shown in the appendix.

Two considerations were important in designing our intervention. First, asking for a large number of reasons has been shown to be counter effective, as participants who have trouble generating new reasons might come to the conclusion that the anchor was right all along (Larrick, 2004). Sanna et al., (2002) showed that requiring two reasons was effective in decreasing hindsight bias, while listing 10 reasons was not. Based on this, six informal consultations were conducted in which the scenarios were presented to academics and public employees. They were asked to state three reasons why the anchor was inappropriate. The rapidness of responses was considered. If the third reason was preceded by a pause, the difficulty of coming up with the third reason was discussed. In general, respondents took more time to generate the third reason and responded by saying that the first two were easy but the third reason was more difficult. Second, the direction in which the anchor was inappropriate was specified, as this has been shown to be more effective than asking people to list reasons in general (Chapman & Johnson, 1999). Earlier research showed that justification is not sufficient (Belardinelli et al., 2018) and that respondents need to allow opposing thoughts in order for the intervention to work.

The experiment has four groups. The first and second groups replicate the experiments reported in Bellé, Cantarelli, and Belardinelli (2017) and Bellé, Cantarelli, and Belardinelli (2018), respectively, to determine the extent to which the biases they revealed exist in a U.K. context. These groups are labelled the *low* and *high anchor* replication groups. Our replications are registered under <https://osf.io/mye2h/> on the Open Science Framework and use the materials of the original authors as well as Brandt et al.'s (2014) replication recipe. The third and fourth groups test the effect of consider-the-opposite approaches to debias anchoring effects. These groups are labelled the *high* and *low consider-the-opposite* groups. The setup is shown in Figure 3.

Figure 3

Experimental Design and Flow



The dependent variables were the maximum number of days in which employees must respond to an email and the performance score on a 0-100 scale. To check that our randomization was working correctly, we included managerial status, gender, industry of employment, educational background, and age in our experiment (Bellé, Cantarelli, and Belardinelli 2017, 2018).

3.3.1 ANALYSIS

First, the high and low replication groups are compared with a *t*-test (testing hypothesis 1). The replication is successful if a significant difference is obtained in the same direction as the original trial. Effect sizes are compared using Borenstein et al. (2009) with a Z-test. Second, again using *t*-tests, the replication groups are compared with the corresponding high and low consider-the-opposite debiasing groups. Debiasing is effective if the mean of the low consider-the-opposite group is significantly higher than the low-anchor group and is obtained (hypothesis 2) and/or if the high consider-the-opposite group report significantly lower means than the high-anchor group (hypothesis 3). Significance levels are established at 0.017 and corrected by the Bonferroni correction per experiment (0.05/3). We also conducted an exploratory sub-group analysis of managers and employees. Data and materials are available on <https://osf.io/mye2h/>.

All effect sizes are calculated using Lakens (2013). The magnitude of the effect sizes is reported in accordance with Sawilowsky (2009). Sawilowsky (2009) expanded the reporting of effect sizes by including reports of very small d (.01), very large (1.2), and extremely large effects (2.0), in addition to effect sizes as developed by Cohen. He did this as a reaction to Cohen's warning about an inflexible approach to effect sizes, leading to Cohen's original values to become standards. Sawilowsky (2009) also wanted to describe more effect sizes as observed in reality.

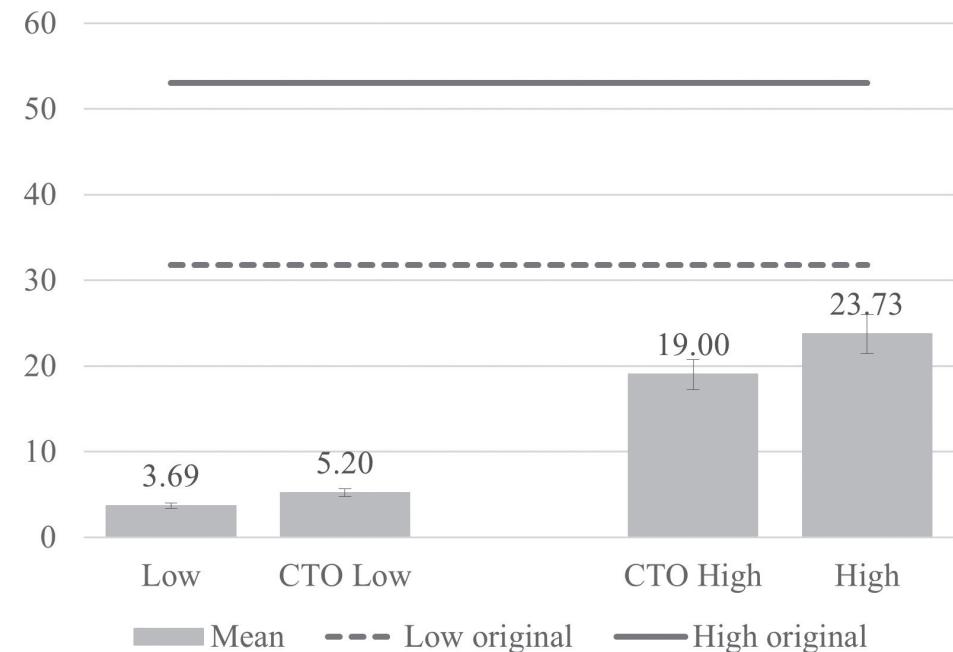
3.4 RESULTS

Our randomization checks showed no differences based on sector, manager, gender, or age. However, the checks did reveal a small difference between educational backgrounds. All the descriptives for each group are shown in Table 5.

The results of experiments 1 and 2 support the first hypothesis, that participants in the high-anchor replication groups will report estimates that are significantly higher than the estimates from participants in the low-anchor replication groups. This indicates that anchors can affect managerial decisions concerning goal setting and performance feedback. In experiment 1, public managers and employees were asked to report the maximum number of days within which a public employee should reply to inquiries from citizens (replicating Bellé, Cantarelli, and Belardinelli, 2018). Here, the replication was successful: the mean score for the low-anchor condition ($M = 3.69$, $SD = 2.98$, $N = 306$) was significantly lower ($t[316] = 16.93$, $p = .00$) than the score for the high-anchor condition ($M = 23.73$, $SD = 20.46$, $N = 305$). However, the effect size differs from the original study. The original study showed a medium effect size (Cohen's $d = 0.41$), whereas in the current study, the effect is very large (Cohen's $d = 1.37$). Furthermore, the mean minima and maxima differ from the original experiment. Bellé, Cantarelli, and Belardinelli (2018) reported a low-anchor mean of 31.82 and a high-anchor mean of 53.07. Our means (3.69, 23.73) indicate a shift of approximately 30 absolute points in the U.K. context. The means of the replication groups of experiment 1 are shown in Figure 4.

Figure 4

Means of Experiment 1 and Bellé, Cantarelli, and Belardinelli (2018) in Days



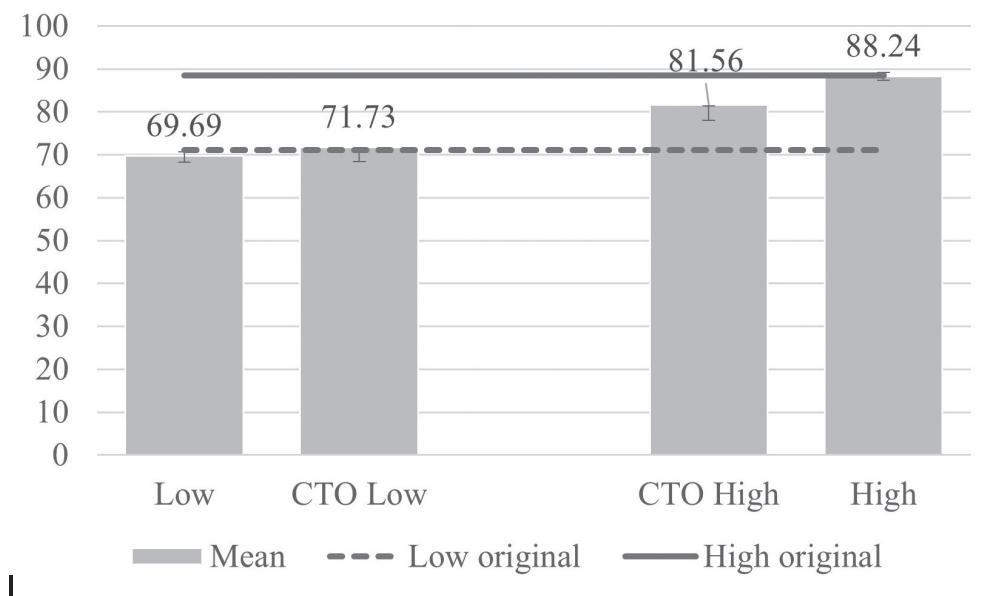
In experiment 2, public managers and employees were asked to rate the performance of an employee (replicating Bellé, Cantarelli, and Belardinelli, 2017). In experiment 2, the replication was also successful: the mean rating for the low-anchor condition ($M = 69.69$, $SD = 10.18$, $N = 306$) was significantly lower ($t[595] = 24.27$, $p = .000$) than the rating for the high-anchor condition ($M = 88.24$, $SD = 8.68$, $N = 306$). In this experiment, the means were very similar to those in the original research in that Bellé, Cantarelli, and Belardinelli (2017) reported a low-anchor mean of 71.07 and a high-anchor mean of 88.47. Nevertheless, the effect size differed. In the original study, the effect size was very large (Cohen's $d = 1.21$). In this study, the effect could be labeled extremely large (Cohen's $d = 1.96$). The Z-tests provide Z-scores of -7.69 for the goal-setting experiment ($p = .000$) and -5.69 for the performance rating experiment ($p = .000$). This means that the effect sizes of the original studies differ significantly of those of the replication studies. The means of the replication groups for experiment 2 are shown in Figure 2.

The second hypothesis, that participants in the low-anchor consider-the-opposite group will report estimates that are higher than participants in the low-anchor replication group, was also corroborated. This means that consider-the-opposite interventions can indeed debias goal-setting and performance feedback practices. Our consider-the-opposite intervention significantly increased ($M = 5.2$, $SD = 4.04$, $N = 298$) ($t[545] = 5.20$, $p = .000$) estimates from the low-anchor group in experiment 1 (low-anchor values $M = 3.69$,

$SD = 2.98$, $N = 306$). The effect size is medium (Cohen's $d = 0.43$). Figure 2 shows the differences between the consider-the-opposite group and the low-anchor group in experiment 1. In experiment 2, our consider-the-opposite intervention significantly increased ($M = 71.73$, $SD = 7.17$, $N = 297$) ($t[549] = 2.85$, $p = .005$) estimates for the low-anchor group (low-anchor values $M = 69.69$, $SD = 10.18$, $N = 306$). This is a small effect size (Cohen's $d = 0.23$). Figure 5 shows the differences between the consider-the-opposite group and the low-anchor group in experiment 2.

Figure 5

Means of Experiment 2 and Bellé, Cantarelli, and Belardinelli (2017) in Performance Ratings



Our results further support hypothesis 3, that participants in the high-anchor consider-the-opposite group will report lower estimates than participants in the high-anchor replication group. The consider-the-opposite intervention produced significantly lower ($t[567] = -3.192$, $p = .001$) estimates ($M = 19.00$, $SD = 15.59$, $N = 295$) than the high-anchor group (high-anchor values $M = 23.73$, $SD = 20.46$, $N = 305$) for experiment 1. This is a small effect size (Cohen's $d = 0.26$). The consider-the-opposite intervention also significantly lowered ($t[600] = -9.31$, $p = .000$) performance ratings ($M = 81.56$, $SD = 8.93$, $N = 296$) compared with the high-anchor group (high-anchor value $M = 88.24$, $SD = 8.68$, $N = 306$) for experiment 2. Here, the effect size is large (Cohen's $d = 0.76$). Figure 3 shows the results of experiment 2.

Table 5
Descriptives and differences by group by condition

Characteristics	Low	CTO Low	CTO High	High	All
% female	80.5	78.4	78.2	78.1	78.8
% manager	32.2	30.2	32.7	29.7	31.2
Average age	39.41	39.34	39.19	38.88	39.21
Sector of employment					
% healthcare	23.5	22.6	25.1	22.5	23.4
% education	41.0	41.0	39.6	42.8	41.1
% administration	15.6	13.4	16.8	13.4	14.8
% other	19.9	23.0	18.5	21.2	20.6
Educational background*					
% technical and scientific degree	22.5	24.3	28.1	32.4	26.8
% social and humanities degree	45.9	48.2	46.9	48.4	47.3

Notes: The differences between groups were tested through chi-square tests apart from the difference in average age, which was calculated using an ANOVA. Significant differences (<0.05) are indicated by an asterisk. CTO = consider-the-opposite.

The sample consists of managers and employees. Even though we did not hypothesize differences between these groups beforehand, we conducted exploratory sub-group analyses. These analyses indicate that managers as well as employees were susceptible to anchors in both scenarios. For the consider-the-opposite intervention, we found that the effect depends on context. More specifically, we found that the consider-the-opposite intervention worked for the high anchor in the performance feedback scenario and the low anchor in the goal-setting scenario for managers. For the high-anchor goal-setting scenario, managers did not significantly adjust their reports while employees did. For the low-anchor performance rating, however, the opposite effect occurred, and managers did significantly adjust their anchors while the employees did not. In tables 10 and 11 in the Appendix D, all results of the exploratory sub-group analyses are shown.

3.5 DISCUSSION

Recent research has shown that decisions by public managers are affected by cognitive biases. However, there is room to strengthen this body of knowledge, and, further, strategies to mitigate these biases are rarely studied. This article replicates two experiments representing two core internal management practices in a distinct institutional context. It shows how anchoring bias, one of the most robust of the biases identified, can be mitigated in the public sector by using a low-cost, low-intensity, and thus scalable, consider-the-opposite strategy. This article has several implications.

CHAPTER 3

First, anchoring bias replicates across institutional contexts in our experiment. This empirically generalizes earlier findings (Bellé, Cantarelli, and Belardinelli, 2017, 2018; Walker, James, and Brewer, 2017). Nevertheless, statistical significance and effect direction only tell a part of the story. Effect sizes are vital when considering replications (Patil et al., 2016). Anchoring effects in the current study are significantly larger compared with the original study. For instance, for the experiment on goal setting, in the United Kingdom, a very large effect was found, compared with a medium effect in the original Italian study (Bellé, Cantarelli, and Belardinelli, 2018). Furthermore, the mean minima and maxima shifted by about 30 absolute points compared with the original study. The differences in effects between Italy and the United Kingdom could be explained by multiple factors, including the survey, timing, and language. Our survey, for instance, exclusively focused on anchoring, while in the original experiments, the scenarios were part of a lengthier survey in which different biases were tested. Political reference levels also could have caused different effects for the goal-setting scenario (Holm, 2017). In other words, existing anchors such as the maximum number of days required by FOIA laws might influence the effect of anchors. Political reference levels could give an indication of the “right” answer, limiting the influence of anchors as a cue of important information on decision-making (Furnham and Boo, 2011). Apart from that, other characteristics of the law could have an effect. The U.K. FOIA law has been in place since 2000, while the Italy law was accepted in 2016. While the law has been relatively successful in the United Kingdom, in Italy, the effects of the FOIA on citizens requests remain contested (Diritto di Sapere, 2017; Worthy, 2010). As such, the age of the laws, the knowledge of the laws and when to apply them, as well as compliance with laws might affect anchoring bias.

Second, our consider-the-opposite strategy achieved the desired outcome in all four cases. Nevertheless, the effect sizes ranged from small to large. This indicates that the effectiveness of consider-the-opposite strategies varies case by case. This research shows that consider-the-opposite is most effective in situations in which a high anchor is presented. In these cases, the anchor could be considered more extreme. Some authors claim that extreme anchors lead to a larger anchoring effect than anchors that are more reasonable (Furnham and Boo, 2011). On top of that, our consider-the-opposite intervention does not fully remove the influence of anchoring, which may imply that anchoring bias is hard to remove completely with debiasing interventions such as the consider-the-opposite strategy. The latter could be explained by different reasons for a person to follow an anchor (Furnham and Boo, 2011). For instance, people could perceive the anchor as being relevant to the problem at hand. Subsequently, consider-the-opposite might only target some of the causes.

Third, our sub-group analyses indicate that anchoring has an effect on both managers and employees. The effect of our consider-the-opposite intervention nevertheless depends on context. For instance, managers adjust goal setting when asked to listed reasons why the low anchor is inappropriate. Managers do not adjust goal setting for the high anchor. On top of that, employees lower their judgment when asked to consider the opposite in the case of a high anchor and move toward a level comparable with the

managers' reports. This might indicate that managers already reflect more critically on the high anchor or have a clearer view of political reference levels, such as FIOA laws (Holm, 2017). The opposite effect occurs for the performance feedback scenarios. In case of a low anchor, employees do not significantly adjust their ratings after consider-the-opposite, while managers do. For the high anchor, both groups adjust their ratings.

Fourth, as our research offers a successful low-cost, low-intensity intervention that can easily be applied in public services, the next step would be to translate this method to real public management practices. A couple of difficulties arise in doing so. First, determining the number of reasons to be provided is crucial, and this might differ in each case. This requires a case-by-case approach. Our general advice is to not require too many reasons as this can backfire (Sanna et al., 2002). Nevertheless, provided an adequate number of reasons is established, a consider-the-opposite approach could possibly be institutionalized in formal and informal ways (Secunda, 2012). This brings us to the second concern: asking managers and employees to formally write down reasons why the anchor is inappropriate might induce a sense of accountability, which could have an effect in itself (Aleksovska et al., 2019).

Fifth, the consider-the-opposite strategy could also be applied to counteract other association-based biases, such as confirmation bias and hindsight bias (Larrick, 2004). Additionally, a consider-the-opposite approach could be adapted to other internal management practices in the public sector such as practices to build trust or increase participation in decision-making (Favero et al., 2016; Pedersen & Stritch, 2018a). Furthermore, possible applications are related to street-level decisions such as during client-employee interactions, where associations concerning client deservingness are known to play a role (Guthrie & Orr, 2006; Jilke & Tummers, 2018; Schafer & Schafer, 2018).

3.6 CONCLUSIONS

This survey experiment investigated whether anchoring bias replicates in decisions representing internal management practices across institutional contexts, the relevance of the anchoring effect among managers and employees, and whether public managers and employees can be debiased by a low-cost, low-intensity consider-the-opposite intervention.

3.6.1 LIMITATIONS

Our research inevitably has limitations. The main limitation of this research is that the experiments focused on control and internal validity, as they were conducted through an online sample based on simplified fictional scenarios (Bouwman & Grimmelikhuijsen, 2016; Harrison et al., 2004). Real-world scenarios might differ and involve more information or more complexity. This could affect anchoring bias and the consider-the-opposite strategy on decision-making. The effect of anchoring bias and consider-the-opposite could, for instance, be smaller in a real-world scenario. Even though these limitations

exist, our scenarios are externally relevant for two reasons. First, anchoring bias proves to be robust across a variety of experimental manipulations and contextual factors (Furnham & Boo, 2011). Even if anchors are self-generated, people do not adjust their estimates sufficiently (Epley & Gilovich, 2001). On top of that, increased understanding of the problem at hand does not discard anchoring bias. Experts are susceptible to anchoring bias, too (Englich et al., 2006; Guthrie & Orr, 2006). Our sub-group analyses also indicate that the anchoring effect is relevant for managers as well as employees. For the consider-the-opposite intervention, we have less empirical knowledge relating to the generalizability of the consider-the-opposite strategy. The sub-group analyses indicate that the effectiveness of the consider-the-opposite strategy depends on context. This sub-group analysis should be interpreted with caution, however, as it is correlational in nature and we lack power to detect small effects (Gerber & Green, 2012).

Another limitation of our research could be the use of the term “bias,” which implies the relationship of a cognitive shortcut to negative outcomes. This has generated criticisms by Gigerenzer and coauthors (Gigerenzer, 1991; Gigerenzer et al., 2008; Gigerenzer & Goldstein, 1996), for instance. The main criticism of Gigerenzer and colleagues is that cognitive biases are tools for humans that help decision-making instead of impairing it (see also Vis 2019). This relates to discussions on what constitutes rationality. In the tradition of Kahneman and Tversky, heuristics can lead to suboptimal decisions compared with a normative standard, oftentimes in line with expected utility theory. In the Gigerenzer tradition, heuristics allow people to make decisions that fit the environment. Although a thorough discussion of rationality relating to heuristics is beyond the scope of this article (discussions include Kahneman and Tversky 1996; Samuels, Stich, and Bishop 2002; Stanovich 2011) two important points of this discussion should be emphasized. First, heuristics are not inherently bad or good, but their merit depends on context (Gigerenzer & Brighton, 2009; Tversky & Kahneman, 1974). Second, the discussion of heuristics is, in the end, a normative discussion that relies on philosophical questions (Hands, 2014). Therefore, in this article, expected utility theory is assumed as a normative ideal. Scholars are invited to use other normative standards to interpret the results of this study.

We draw three main conclusions and propose directions for future research. First, anchoring is relevant across institutional contexts. Future research could explore whether there is an interaction with the effect of institutions, or more specifically rules and expectations, on the effect of cognitive biases. Second, anchoring bias has an effect on both managers and employees, as such being a manager does not remove anchoring bias. Third, a consider-the-opposite strategy can mitigate anchoring effects in our goal-setting and performance feedback scenarios. This strategy consists of requesting two reasons why the anchor is inappropriate. Effectiveness, however, depends on at least on two contextual factors. First, effects seem stronger when anchors are more extreme. Future research could also specify to what extent the perception of extremeness matter, for instance by asking whether respondents find specific anchors too low or too high. Second, managers might react differently to the consider-the-opposite strategy than employees. Managers for instance could have more knowledge of relevant laws. On the

other hand, managers can be debiased in other cases. Future research could focus solely on managers for the aforementioned scenarios. This strategy has the potential to be used to address other association-based biases, as well as with other internal management practices and other types of employees such as street-level bureaucrats.

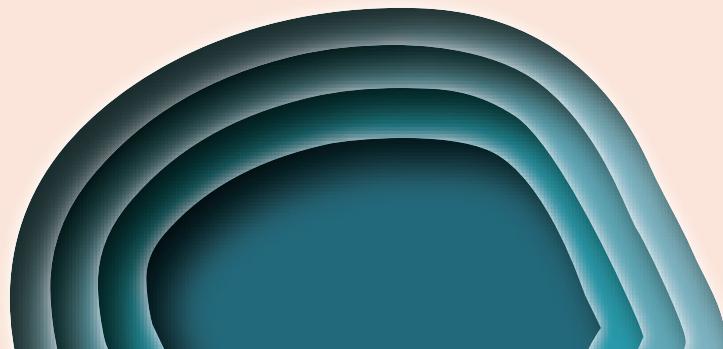
The practical implications of this research are that anchoring should be taken seriously in public management contexts, and could influence goal-setting and performance feedback practices. Therefore, it should be considered when designing jobs and tasks. Apart from that, our research indicates that in cases in which anchoring bias is a problem, a consider-the-opposite strategy is a promising tool to mitigate anchoring bias. This research should be seen as a first step toward mitigating anchoring bias. The next step is to test it in real-world scenarios. As this research has shown that anchoring bias can work differently depending on context and a consider-the-opposite technique mitigates anchoring bias in a controlled setting, we encourage scholars to test these results in more realistic settings. Future research should therefore focus on real-world scenarios prone to biases and field experiments to test the consider-the-opposite strategy.



CHAPTER 4

**Don't speak: exploring how personal
and contextual factors affect politicians'
participation in online deliberations**

This chapter is based on the following manuscript: Nagtegaal, R. Don't speak:
exploring how personal and contextual factors affect politicians' participation in online
deliberations. *Under review.*



ABSTRACT

The COVID-19 crisis caused public decision-making bodies to have to deliberate online. Online environments are known to lead to inequalities because of differences in access, and in the way people actively participate online. We know that personal and contextual behavioral factors could explain these differences, while simultaneously offering possibilities for intervention. Therefore, we ask the following research question: *How and why do personal and contextual factors influence active participation in online deliberation through videoconferencing by politicians?* 15 semi-structured interviews were conducted with council clerks and local council members. This data was coded into the Theoretical Domains Framework - a comprehensive behavioral framework. Our results show that personal factors (for instance, beliefs about whether responding serves a purpose), and contextual factors (such as rules in meetings) influence active participation. We recommend practitioners to offer organizational support to address a lack of skills, intervene upon beliefs about consequences and make fitting rules by sharing best practices.

4.1 INTRODUCTION

In 2020 the world went dramatically online in a way not seen before. The COVID-19 crisis forced us out of our offices and into our homes, where we interacted through the blue light of our screens. However, this did not just impact our watercooler chats. Political decision-making bodies, such as municipality councils, supra organizational networks and councils of ministers also went online. These public decision-making bodies had to continue making essential decisions, including how to manage COVID-19 and its effects and how to deal with ongoing political issues about dividing resources and investing in the future.

Ideally, these decisions would be based on deliberation, about carefully weighing arguments and achieving the best solution. In a representative democracy, deliberation happens through representatives of a group within society participating within a deliberation. Nevertheless, ideal deliberation is often hard to achieve, might even be harder to achieve when we go online (Bächtiger et al. 2010; Strandberg and Grönlund 2018). This happens because people might use Information and Communication Technologies (ICTs) differently, as research on the digital divide, which refers to inequalities caused by their access, skills and use of ICTs explains (Scheerder et al., 2017; Van Deursen & Van Dijk, 2019).

We are interested to research if differences in active participation occur when deliberations of public decision-making bodies go online. Active participation means the extent to which people voice their opinions in deliberative contexts (Albrecht 2006). Because of the novelty of formal deliberative bodies deliberating online, we explore what causes possible individual differences. We do this by studying behavioral factors that influence active participation. Behavioral factors are factors that build on psychology and that influence individual behaviors. These can be personal, such as skills, but also context related, such as social pressures (Michie et al., 2014). Thus, we are interested in the following research question:

How and why do personal and contextual factors influence active participation in online deliberation through videoconferencing by politicians?

This question is answered by conducting a qualitative study based on semi-structured interviews. As a case, we chose online deliberation through videoconferencing of local council members in the Netherlands. This case was chosen because we expected to find some variability between council members as well as municipalities and issues have been detected in this context (Peters et al., 2020). The study explores personal and contextual factors influence behavior by analyzing data using a comprehensive behavioral model: the Theoretical Domains Framework (Cane et al., 2012; Michie et al., 2005). By doing this we provide an overview of the range of factors that influence active participation by politicians in online deliberation.

This research contributes to literature and practice in three ways. First, it contributes to online deliberation research. We focus on specifically studying online deliberation between politicians. Other online deliberation research has to date mainly focused on including citizens, who were previously unheard, in deliberation with politicians or other citizens (Friess & Eilders, 2015; Strandberg & Grönlund, 2018). Yet, most decision making between politicians continued to happen in face-to-face settings such as meeting rooms. Online deliberation through videoconferencing between public decision-making bodies has emerged recently, as a result of the COVID-19 crisis. Previously, this was not common practice. We also study real-time online deliberation through videoconferencing, which is to our knowledge not researched yet. Most research focusses on asynchronous communication such as through online fora or synchronous communication through chat (Friess & Eilders, 2015; Strandberg & Grönlund, 2018).

Second, this article contributes to behavioral public administration research by showing how we can study processes beyond citizen and public employees behavior as well as beyond biases by using an comprehensive model of behavioral factors (Hallsworth et al., 2018). In addition, we use a qualitative method, which is rare in behavioral public administration research (Bhanot & Linos, 2020). This allows us to focus on real issues reported by the politicians rather than fictional scenarios. Also, instead of focusing on one topic, the approach considers the issue from a more comprehensive and contextualized perspective (Roberts, 2020).

Third, for practice, the study conducts a behavioral analysis identifying how behavioral factors matter for active participation and how these can be used to initiate behavioral change by intervening upon them (Michie et al., 2014). This is important as politicians are assumed to, to some extent, represent interests of their voters (Brown, 2018). Intervening might thus be essential to safeguard the democratic character of formal deliberations (Moscrop & Warren, 2016).

4.2 THEORY

4.2.1 ACTIVE PARTICIPATION IN ONLINE DELIBERATION

This paper concentrates on finding differences in active participation between politicians during online deliberations. Deliberative democracy and representative democracy are sometimes understood as opposites due to the actors (respectively citizens or politicians) involved in the decision-making process (Landemore, 2017). Yet, the practice of 'trial by discussion' is one of the institutional pillars at the core of representative democracies. This practice can be equated with the practice of deliberation, which is proposed by deliberative democracy theorists. The deliberative democracy movement started out from a normative ideal of democracy (Bächtiger et al. 2018). This ideal democracy was envisioned as having deliberation at the core of it. People in political decision-making processes would listen to each other's opinions, reflect on them and be willing to change theirs. In other words, the core of deliberative democracy is that decisions are made on the basis of deliberation,

of talking and listening to each other, rather than just on an aggregation of fixed preferences. Conditions for this were that deliberation was supposed to happen in an environment of mutual respect, equal recognition, reciprocity and sufficient equalities in power. This is related to who gets heard in deliberative settings (Brown, 2018). Although deliberation rarely happens under these ideal settings, deliberative practices are part of representative democracies in governmental organizations, such as municipalities. In such bodies, formal deliberation is part of the decision-making process, while being shaped and accompanied by institutions (Fairclough, 2019; Landemore, 2017).

What makes deliberation democratic is equality in opportunities for participation (Moscrop & Warren, 2016). Exclusion in deliberation occurs when people do not have the same resources to participate as others in the political process (Beauvais & Beauvais, 2018). Since the rise of Information and Communication Technology (ICT), new possibilities for participating in deliberation in online settings have emerged (Strandberg & Grönlund, 2018). Online deliberation refers to all kinds of discussion online, for instance, on social media, such as Facebook or Twitter or online fora (Friess & Eilders, 2015). This development has been especially interesting since the virtual sphere offers new opportunities to include citizens in the political decision-making process (Dekker & Bekkers, 2015). Now, for the first time, many formal bodies are also deliberating online through videoconferencing. Yet, we do not know how politicians behave differently in online deliberative settings.

The use of ICT can create inequality, which is described by the digital divide movement (Van Deursen and Van Dijk, 2019). First wave digital divide scholars focused on inequalities caused by differences in internet access. However, participation goes beyond having access and beyond being formally present in a meeting. It extends to people actively participating within this setting. As Young (2000) explains, some participants may be marginalized during deliberation if they have lower chances to be heard, introduce topics, make contributions, or suggest or criticize proposals. This is visible in digital divide literature as well. Second and third level digital divide scholars emphasized that inequalities can occur because of differences in use and skills (van Deursen and Helsper, 2015). In other words, we can assume that online environments might create some problems for representation, because people might differ in how they participate in online environments. In the case of elected officials, elected officials can be assumed to, to some extent, represent interests of their voters (Brown, 2018). This paper thus focusses on individual differences in actively participating in online environments (Albrecht, 2006), where politicians voice their opinions in online deliberations.

4.2.2 EXPLORING PERSONAL AND CONTEXTUAL FACTORS

This section explores which personal and contextual factors can explain differences between politicians by studying behavioral factors. Behavioral factors refer to those aspects that can explain why individuals display behaviors. As this occurs at the individual, at the micro-level, these factors are often psychological (Atkins et al., 2020). Because

of the novelty of the situation, we do not know yet which behavioral factors might be relevant. Recent research connecting behavioral factors and deliberation indicates that this could be a variety of factors. The BIT UK has for instance studied deliberation from a behavioral perspective, focusing on bias (Hallsworth et al., 2018). They state that the way people interact with each other in groups can breed biases. The digital divide literature shows that former use and skills matter and online deliberation research also points to individual factors such as such as skills relating to technology use and motivation to participate (Scheerder et al., 2017; Strandberg & Grönlund, 2018).

We chose to use a comprehensive behavioral model to explore what influences participation, given the variety of behavioral factors involved. Different models exists that bundle theories of behavior such as the Consolidated Framework for Implementation Research (CFIR) and the Knowledge to Action framework (Birken et al., 2017). This research employed the Theoretical Domains Framework (TDF). The TDF was developed to bundle and simplify a plethora of different theories on behavioral change (Michie et al. 2005). These theories on behavioral change can be used to explain why people behave the way they do. The authors identified 33 theories and 128 causes that explain behavioral change and integrated these into one single framework.

The TDF identifies 14 different factors that influence behavior. These factors can be related to the person, such as knowledge, skills, memory, attention and decision-making processes, intentions, goals and beliefs about consequences. An example, related to online participation, is that people might lack the technical skills needed to participate. Dimensions can also be related to the context: to the environment and social influences. An example in our research is politicians pressuring other politicians to speak up. Appendix F, lists all the theoretical domains, the definitions and possible relevant constructs.

To our knowledge, the TDF has not been applied to political deliberation as yet. The TDF offers a comprehensive tool to study participation from a micro perspective. The TDF is especially useful because apart from personal factors, such as skills and motivation, the model also includes context dimensions such as social relations and support, as well as group and organizational factors such as resources and leadership (Nilsen & Bernhardsson, 2019). We expect these factors to also be important especially when studying deliberation, as deliberation happens by individuals acting in and interacting with a groups (political parties) and organisations (governmental organisations) (Barabas, 2004; Strandberg & Grönlund, 2018). Although the TDF offers a comprehensive, theory-based model, the downside is that coding can be time consuming, the interpretation of categories is difficult and categories can overlap (Phillips et al., 2015). The TDF also does not show linkages between different constructs. Section 4.5 below discusses the limitations of using the TDF.

4.3 METHODOLOGY

This section describes the case studied, the operationalization of the variables, and the procedure and analysis. We research active participation in online deliberation by studying Dutch city councils. We chose this case for three reasons. First, local council members are known to be quite diverse in terms of age and education (Overheid in Nederland, 2018). Because, for instance, a factor like age is related to less experience with technology (Niehaves & Plattfaut, 2014), we expect to find variation. Second, because many different municipalities went online, we can also explore the effect of various organizational and group factors. Third, we have indications that active participation in online deliberation might be an issue. In a policy evaluation Peters et al. (2020) mentioned that less interaction occurred because of digital council meetings. Below, we describe the case more in depth.

4.3.1 CASE: ONLINE DELIBERATION IN DUTCH LOCAL POLITICS

The Netherlands has three traditional governmental layers. Local government is closest to civilians and consists of municipalities, where city councils are the most powerful organs (Breeman et al., 2016). A city council is the governing political body, the local parliament, charged with controlling and regulating of daily governing of the municipality. In the Netherlands 352 municipalities existed on the 1st of january 2021 (Centraal Bureau voor de Statistiek, 2021).

On the 7th of April 2020 the Dutch Senate approved a law that allowed digital meetings and digital decision-making for city councils and other local political bodies (Binnenlands Bestuur, 2020). The law is still in place at the moment of writing. City councils comprise of between 9 to 45 councilmembers, depending on the size of the municipality (Nederlandse Vereniging voor Raadsleden, 2020b). City councils convene bi-monthly or monthly in public city council meetings, which are open to the public and where political decisions are being made. The alderman chairs these meetings. Apart from these 'general' council meetings, other council meetings exist.

City councils have different meeting structures. Some city councils have committee councils dealing with different content. For instance, the city of Amsterdam has a committee on 'living and building', but also a committee on 'economy and financial affairs' (Gemeente Amsterdam, 2020). Meetings can also have different types of goals. In the Netherlands three main goals are distinguished: forming an image; forming an opinion; and, making a decision (in Dutch, BOB-model) (Boers, 2017).

In our study, we interviewed council members and council clerks. City councils are supported by council clerks (Nederlandse Vereniging voor Raadsleden, 2020a). A council clerk can support local council members in at least three ways. First, as secretary to make sure meetings are held and documented properly. Second, as an intermediary between the local council and other relevant parties, such as the media. Third, to offer substantive support to council members, for instance, by providing technical information.

4.3.2 OPERATIONALIZATION

We study how behavioral factors affect active participation. We elaborate on how we operationalize these variables below.

Behavioral factors: The focus is on factors present in the Theoretical Domains Framework (TDF) (Cane et al., 2012; Michie et al., 2005). The TDF has fourteen different factors, referring to personal factors including skills, beliefs about capabilities and goals, but also contextual factors, for example social pressures and environmental aspects such as resources. An example of a behavioral factor would be perceptions about the consequences of actively participating. In the context of this study this can mean that a politician states that actively participating is essential to ensure that the decision-making process is deliberative. Another example would be when people experience pressure from peers to speak up. The appendix gives an overview of all the factors and their definitions.

Active participation in online deliberations: We focus on active participation as an outcome, by concentrating on the activity of contributing to deliberation by voicing opinions (Albrecht, 2006). In this context, this could happen through the giving of an original contribution, but also by responding to another person's contribution. The latter is referred to as an 'interruption' in the Dutch local government context. We also study if a person is able to get the word by the chair. We do not study access or to what extent the voice is taken into account by others. We also do not study the quality of the participation, or more in general, the quality of the deliberation.

4.3.3 PARTICIPANTS AND RECRUITMENT

Interviews were conducted with municipality staff and council members. We aimed for 12 interviews to reach saturation as suggested by Guest, Bunce, and Johnson (1995). We used purposeful sampling to contact people from different municipalities who had reported different experiences with digital council meetings in the media. Snowball sampling was used, by asking participants if they could help us get to other participants. Here, we asked to provide contacts which were diverse. We approached people through e-mail or by phone. We aimed to interview multiple people within different municipalities to account for unknown within municipality variation. An overview of respondents is shown below in table 6.

4.3.4 PROCEDURE AND ANALYSIS

We reached saturation by conducting 15 semi-structured interviews on the phone and through videoconferencing software. At the beginning of each interview, we informed participants of the topic of the interview, how the data was going to be used and stored, and asked for their consent in using the data and recording the interview. All participants consented. Interviews lasted from 30 minutes to 2.5 hours and, sometimes, were conducted in multiple sessions. Interviews were semi-structured. The topic-list is given in

Appendix E. Participants were given the space to talk about topics of their choosing. All interviews were transcribed.

Table 6*Overview of the respondents*

Respondent code	Municipality (code)	Job
R1G	M1	council clerk
R2R	M1	council member
R3R	M1	council member
R4G	M2	council clerk
R5G	M3	council clerk
R6G	M4	council clerk
R7R	M2	council member
R8R	M1	council member
R9R	M4	council member
R10R	M5	council member
R11G	M5	council clerk
R12G	M6	council clerk
R14R	M5	council member
R15G	M7	council clerk
R16R	M6	council member

We used a combination of deductive and inductive coding. This means that we coded sections referring to active participation and a cause by placing the cause in the Theoretical Domains Framework (TDF). We left room for induction due to the exploratory nature of this study (Mcgowan et al., 2020). For instance, if factors fell under the TDF, but were psychological, we also planned to code these. In accordance with Atkins et al. (2017) only coded text relevant to the behavior of interest in this study. We, moreover, coded text that could fall into multiple categories in all of these categories. Afterwards, we constructed a conceptual model to show linkages between different constructs.

4.4 RESULTS

This section highlights our most important findings. First, we describe what active participation in deliberations through videoconferencing by local council members looks like. We then go into personal and contextual behavioral factors affecting active participation. We end by presenting a conceptual model.

4.4.1 ACTIVE PARTICIPATION IN DELIBERATION THROUGH VIDEOCONFERENCING BY LOCAL COUNCIL MEMBERS

For clarity we would like to give a description of the context first. From the 4th of April 2020 until the moment of writing local municipalities were allowed to conduct decision-making meetings online. This led to different approaches. Some municipalities went fully online, others partially held meetings online and some did not conduct meetings online at all. The municipalities that went online used videoconferencing software such as Zoom, Teams and Starleaf. These normally allow a limited number of people to be visible on screen and provide a chat function as well as the possibility to digitally raise a hand. Respondents report difficulties in using the software. For example, respondents find it *hard to keep an overview* when many things are happening at once. This also occurs because people were using different devices (such as a tablet or phone) at the same time. People also report having to focus more and some noted that digital meetings were tiring. *Technical difficulties* are often reported as well, and these were sometimes related to the resources people had at home.

In this article, we are interested in active participation. Active participation can take on different forms. One can actively participate through making an original contribution but also through responding to another person. This means that no active participation occurs when people do not voice their opinions in council meetings. Active participation can also happen through *non-verbal communication*. In this sense, people can react by making sounds (for instance laughing, sighing) or by body language (for instance frowning, smiling, nodding). However, through the technology, non-verbal participation in general becomes more difficult, while in other cases non-verbal communication can be excessively visible. Apart from this general sentiment, personal and contextual behavioral factors affect active participation within online deliberations. We go into this below.

4.4.2 PERSONAL FACTORS LEADING TO ACTIVE PARTICIPATION

The results show that different personal factors affect active participation, including skills, beliefs about whether responding serves a purpose, goals, and the power position of politicians.

Skills on dealing with technology and installed rules are sometimes a barrier to dealing with the software, and subsequently, for communicating arguments during the deliberation. Respondent R16R, for example, stated:

R: And with Teams it is the same. That is very hard, for me at least, because I have to press a button and I have to do something in... and by the time that I have been able to do that the meeting has almost ended.

Skills can also refer to the knowledge about 'being on screen'. This means that a lack of skills can result in unwanted non-verbal communication. Someone might not notice that they are very visible. The ability to handle digital meetings is dependent on earlier

skills, which are often dependent on having an interest in technology and one's age. One respondent, R12G, for instance explained:

R: We have everybody and all age categories with every background and level of education. That means a lot for many people. An application like Skype was for some council members something they have talked about once with the grandchildren, but now they have to use it themselves in a council meeting, at least a variant on it. That meant a lot of individual support as well.

These skills can be learnt, although this remained difficult for some council members. They need prolonged support, as witnessed in the previous quote. Some council members also lack motivation to learn the software.

Responding to each other is connected to *beliefs about if responding serves a purpose*, either in bringing a point across, achieving a change of opinion or not 'disrupting' the meeting. For example, if an amendment does not seem to have a majority, it is not viewed as useful anymore to respond. Respondent R7R states:

R: Yes, the decision-making is about motions and amendments, and you will soon think: 'Well I see that has no majority anyways, never mind.' Or 'Putting it together has no purpose' or 'Do you want to have an extra adjournment?'. You see that council members realize faster: 'Well, this did not make it anyways. I can discuss digitally for a very long time but it does not add anything'.

This is connected to experiencing barriers to participate. Respondent R7R explains:

R: That people really wanted to make a point, but, because there was a slightly higher threshold for getting the floor and actually saying something, for a lot of those points that people apparently found less important, they just passed. 'Well, I'm not going to go to all that trouble just to say this or this.' And I did have that feeling. While in a physical meeting I think people would interrupt more often and try to make a point more often.

Council members also can have *goals* to actively participate during digital council meetings on important topics. This goal setting can overcome barriers that are experienced for active participation in council meetings. Respondent R12G explains:

R: but they are going to use self-censorship somewhat selectively, as points become more important. The degree of division over documents in the council meeting is really important for the way it is used.

For group factors, the *power position* which politicians occupy matters for their participation in meetings. One respondent mentioned that he does not get the same opportunities to access or speak in digital meetings compared with others whose opinions are more

common. Another respondent mentioned that she does not have a lot of power and does not feel heard, so does not feel motivated to speak.

4.4.3 CONTEXTUAL FACTORS LEADING TO ACTIVE PARTICIPATION

Apart from personal factors, contextual factors influence online deliberation through video conferencing. These factors are organizational, such as support and rules, but also involve characteristics of the group, such as group size and polarization.

First, the *organizational support* that council members receive varies across municipalities and seems to be related to resources. For example, in one municipality where people were positive about digital council meetings, a lot of resources were used. The support staff put a lot of time and energy into supporting council members, hired an external company to make instructional videos, had multiple people fulfil different chairing roles to not miss anything and offered prolonged support for those having trouble in later stages. Support for council members is important to ensure general skills, which then assist the ability to participate in council meetings.

Second, in council meetings *rules* are often seen to have an important influence on participation. These rules can relate to how people respond to each other in the digital environment, but also to how long contributions have to be. Rules are set by support staff and sometimes the presidium. Rules are seen as useful to keep control of meetings, but also are recognized as a barrier for the participation. Rules, for instance, do not allow people to respond directly, and following rules can delay or lengthen contributions, thereby leading to people not contribute. This does not seem to be dependent on the type of person. One respondent (R9R) mentioned that these rules might even matter more than digital environments:

R: 'Hey you have to report'. 'You have to do this'. 'You have to keep it short', 'You should actually send in your questions in advance', but which (rules) are not useful for a political discussion. You actually kill it with that. I think that's a pity. And that has nothing to do with digital meetings, it has everything to do with how we try to get it right and then decide on rules. We've had some physical meetings in between with the same kind of rules, they've been killed as well. There was no political debate there either.

In one case, rules are seen as positive. In that case support staff chose a topic based deliberation structure for complex topics. This provides the ability to have an overview and focus and, thus, diminishes the possible negative effects of the software. Another example is rules about 'asking for the floor'. In one municipality, council members were asked to post in the chat if they wanted to respond, to be allowed to talk or ask a question. The chair then gave people who wanted to respond the opportunity to speak immediately. In other municipalities, when people raise their digital hand, it is unclear what they wanted. This often lead to them getting the opportunity to speak only after the argument. Two respondents explain:

R5G: What I really like about that chat is that it is immediately clear what you want. Look, if you raise a hand, I don't know if you want to interrupt, or have a question, or want to speak. But because if the chairman sees that someone wants to speak during someone's speech, he does not say: "shut up, because I have someone who raises a hand". But then he says: "go on with your argument". And if he sees in that chat: someone has an interruption with the "ir", then he says: "keep your mouth shut, because I have an interruption. Wait a moment before you continue". And then it makes sense that you have those interruptions immediately, because you should not have them after an entire speech. I would never want that hand as a condition, I want to have that chat function in the screen where they can do that: interruption, word or question mark so to speak.

And R10R:

I: How did it go, the interrupting?

R: We had agreed before the order of the meeting that you would do that after the spokesperson. Yes, of course that's does not work.

I: No. So, it was delayed?

R: Then you're out of momentum. Well you know it didn't get delayed, then we also found out it almost didn't happen.

I: Oh! Okay.

R: Because then it is, if someone gives a five-minute speech and you want to interrupt with a comment in minute one and a half.. Then it is too little, too late.

In that way, small rules can make a big difference for the feedback loop of deliberation. Third, *group factors* are important for deliberation. For instance, *group size* can exacerbate the lack of overview, which is already a problem in the online environment. What also is reported frequently is when *group opinions are polarized*, and sometimes topics are emotional, deliberation is more desired and simultaneously more difficult. This happens because it is harder to keep an overview and because responding to each other is a stepwise process. Respondents R4G:

I: Do you see that in debates about something that is complex?

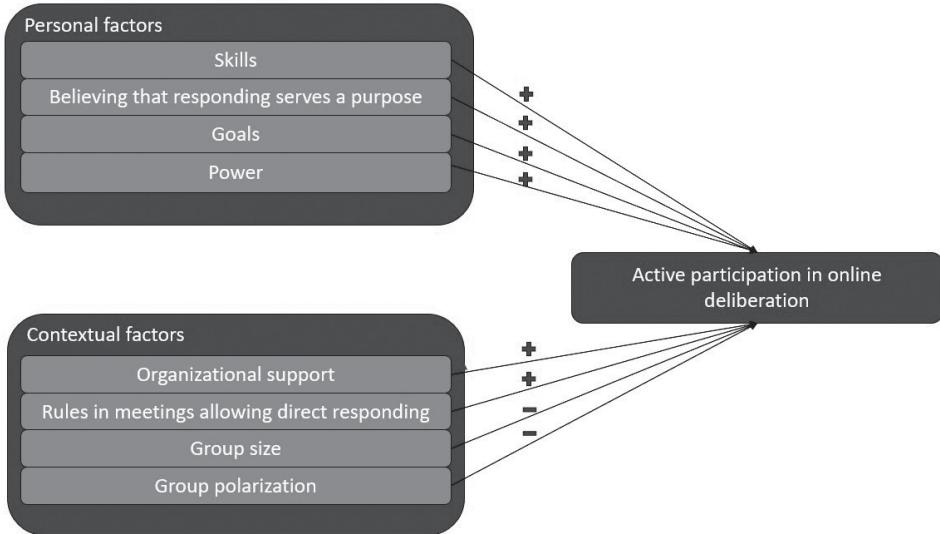
R: Yes, in the digital environment you see in our case many hands being raised. Hands in the computer system. Yes, then you don't know any more who wants to react on someone, or if they have their own contribution. While, what I said, in a physical environment you know instantly because people just point at each other.

Respondent R1G:

R: The moment it gets heated, and people raise their arms at the same time, digitally or not, you have to pay a lot of attention. That gets harder digitally. It is possible, but it goes very step-by-step. At that moment you are losing some of the fun in the debate. I am also... It has its limitations.

From the data above, we generate the following model of behavioral factors influencing participation in online deliberations:

Figure 6
Conceptual model based on the qualitative data



4.5 DISCUSSION

The COVID-19 crisis induced many public decision-making bodies to introduce online deliberation through videoconferencing. Even though the software used was similar, people reported differences experiences in deliberation. Possible explanations can be found in studying behavioral factors. Therefore, we asked the following research question:

How and why do personal and contextual factors influence active participation in online deliberation through videoconferencing by politicians?

First, personal factors such as skills with technology, as well as beliefs about whether responding serves a purpose and goals affect active participation during videoconferencing of public decision-making bodies. These skills depend on one's interest in technology, which is often related to age. Likewise, the power position of a politician might influence these factors. An example is when person without power might think their opinion will be ignored anyway. This implies that some groups of politicians, such as older politicians or people that do not have a lot of power, might actively participate less. Thus, online videoconferencing can disproportionately affect already disempowered groups. Organizational support for these groups can reduce these differences by intervening on skills. This is essential, as the democratic aspect of deliberations might suffer if the needs of these groups are being ignored (Moscrop & Warren, 2016).

Second, various contextual factors affect active participation, including organizational support, rules in meetings, group polarization and group size. Organizational support can reduce differences in skills by intervening on them. As a reaction to the downsides of digital conferencing rules are made, that do not always foster active participation. They can for instance delay responses, making people believe that responding is 'too little too late'. This can affect participation not only of certain individuals, but of all individuals present in the online deliberation. Although rules have been identified as an important factor in earlier work (Friess & Eilders, 2015), organizational factors and group factors seem to be neglected in literature (Strandberg & Grönlund, 2018). The reason for this could be that these factors are especially important in this type of online deliberation which happens within an institutionalized setting and through videoconferencing.

Additionally, the technological environment has some general downsides which are the lack of overview, technical difficulties and non-verbal communication. For the latter, we see that active participation is more than just a verbal endeavor. Earlier online deliberation research showed that social cues are often missing in online environments (Albrecht, 2006). A videoconference context does not solve this problem. In fact, respondents note that non-verbal communication is often lacking and is simultaneously essential for active participation in formal deliberations. This might mean that videoconferencing can be unfit for certain types of deliberations involving big and polarized groups. For instance, bigger groups often mean more people on screen, and subsequently, less attention for individual non-verbal communication. This might mean that videoconferencing can be unfit for deliberations occurring in big groups.

Practically, we make three recommendations. First, prolonged organizational support is needed to assist people lacking skills, especially older people, people without much technical experience and people who are in an disempowered position. This support requires financial and staff resources. Second, municipalities can intervene in beliefs about consequences and goal setting. Beliefs in consequences can, for example, be intervened upon by giving information about what voters think about politicians not actively participating. In table 7, we present some additional behavioral change techniques. Third, setting the right rules is essential for participation, not only of certain groups, but of all politicians. Rule setting needs to be discussed at an intra-organizational level and best practices have to be shared.

This study has several limitations. First, the study is based on interview data which consists of perceptions on behavioral factors and active participation. This means that subjectivity plays a role in the data gathered and causality cannot be established. However, this method also has benefits as behavioral factors, such as motivation, skills and pre-defined rules, are often not visible in the deliberation itself. This means that interview data can be a valuable addition to existing techniques such as focusing on (textual) analysis of deliberation, for instance, by measuring deliberation quality (Steenbergen et al., 2003). Second, although the Theoretical Domains Framework (TDF) provides a comprehensive model of behavioral factors, interpretation into categories was

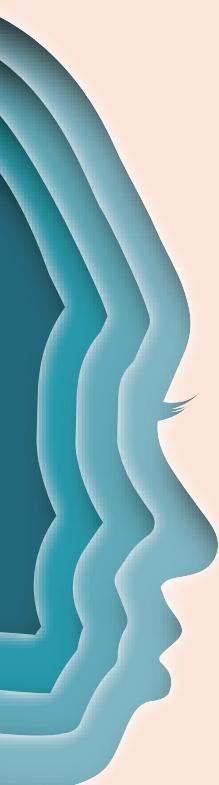
difficult. This is a known downside of using this model (Phillips et al., 2015). Nevertheless, because of the exploratory nature of this study, using this model offered a searchlight to spot possible relevant factors. Third, this study focuses merely on active participation by elected officials in a formalized setting. Complex concepts, such as representation and deliberation cannot be fully captured by focusing on just one aspect (Brown, 2018). Future research can concentrate on other aspects of video-conferencing, such as the inclusion of citizens in these meetings, as well as the content of deliberations and informal practices surrounding online meetings.

Table 7

Possible interventions on beliefs in consequences and goals (based on Michie et al. 2014)

Interventions for beliefs in consequences	Interventions for goals
<p><i>Information about social and environmental consequences:</i> Give information about what the voters believe about politicians not actively participating in online deliberation</p>	<p><i>Goal setting (behavior):</i> Set a goal of a behavior that has to be achieved. For instance, for an important topic a politician might have a goal to raise two points that are important to them.</p>
<p><i>Anticipated regret:</i> Raise awareness about future regret if people do not participate by asking a person if they will regret not contributing in the meetings.</p>	<p><i>Action planning:</i> Undertake detailed planning of performance of the behavior which includes context and frequency. For instance, have someone plan when to respond.</p>
<p><i>Pros and cons:</i> Advise the person to identify the reasons for wanting to actively participate and not wanting to by listing them</p>	<p><i>Review of outcome goal(s):</i> Review if the goal has been met and if goals or behaviors need to change accordingly. For instance, organizational staff might reflect with a person on how much they participated.</p>

We conclude that online videoconferencing between politicians has downsides. Behavioral factors, whether personal or contextual, influence active participation. In order to keep the process deliberative and democratic, we need intervene on these behavioral factors, for instance by discussing rules on an intra-organizational level, or by action planning. This is especially important when groups are big and polarized.



CHAPTER 5

The impact of using algorithms for
managerial decisions on public employees'
procedural justice

This chapter is based on the following published article: Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, 38(1), 101536.

ABSTRACT

Algorithms are used in public management decisions, for instance, to allocate police staff to potential crime scenes. We study how the usage of algorithms for managerial decisions affects procedural justice as reported by public employees. We argue that some public management practices may be more suitable for algorithmic decision-making than others. We hypothesize that employees' perceptions differ depending on the complexity of the practice at hand. We test this through two survey experiments on 109 Dutch public employees and 126 public employees from the UK. Our results show that when a decision is made by an algorithm for practices that are low in complexity, procedural justice increases. Our results also show that, for practices that are high in complexity, decisions involving a public manager are perceived as higher in procedural justice compared to decisions that were made automatically by computers using algorithms. Nevertheless, adding an algorithm to a public manager's decision-making process can increase procedural justice for high complexity practices. We conclude that managers should explore automation opportunities for low complexity practices, but to be cautious when using algorithms to replace public managers' decisions for high complexity practices. In the latter case, transparency about algorithms and open dialogues on perceptions could be beneficial, but this should not be seen as a panacea.

5.1 INTRODUCTION

The idea that data can be used to improve decision-making processes in organizations has become more popular (Anastasopoulos & Whitford, 2019; Desouza & Jacob, 2017). At the same time, technological developments have allowed more and novel applications of algorithms to be involved in human decision-making processes (Burton et al., 2019; Veale & Brass, 2019). On top of that, algorithms have recently been moved up higher in the hierarchy and are becoming decision-making partners or substitutes at the level of leadership (Wesche & Sonderegger, 2019). In other words, algorithms are increasingly being used for managerial decision-making. For instance, some companies, such as Uber, are almost fully substituting managers by algorithms. Other examples are that personalized nudges based on algorithms have been implemented within organizations to change employees' behavior (The New York Times, 2018), while data mining has been used for the selection and evaluation of employees (Strohmeier & Piazza, 2013).

Novel utilizations of algorithms are also used for managerial decisions in the public sector. Examples include: the calculating of optimal routes for collecting municipal waste (Karadimas et al., 2007); analyzing which buildings are more likely to catch on fire to guide which fire safety inspections should be prioritized (Engin & Treleaven, 2019); estimating where the chance of criminal behavior is the highest, and subsequently, send police staff to these so called 'hotspots' (van Zoonen, 2016); the evaluation of teachers' performance (Diakopoulos, 2014; O'Neill, 2016); and, guiding physicians behavior through nudges based on algorithms in healthcare (Nagtegaal et al., 2019).

Algorithmic decision-making is, however, far from uncontested (Veale & Brass, 2019; Zarsky, 2016). The debate on the value of algorithms focusses on multiple aspects, including accuracy, power and bias. People can, moreover, display algorithm aversion – which is a tendency to prefer human decision makers over algorithmic ones (Burton et al., 2019). In this paper, we focus on the effect of including algorithms in managerial decisions on procedural justice perceptions as reported by public employees. Procedural justice refers to the extent that the process of decision-making is perceived as being fair (Colquitt, 2001; Lind & Tyler, 1988). Procedural justice contributes to perceptions of legitimacy (Mazerolle et al., 2013). Algorithmic decision-making has been identified as a problem for the legitimacy of decision-making processes (Danaher, 2016) as they are often opaque and might introduce bias (Janssen & Kuk, 2016).

We expect that perceptions of procedural judgment differ depending on the involvement of the public manager and the algorithm, as well as the complexity of the practice at hand. Building on the work of Zouridis et al., (2020), we distinguish three categories of algorithmic public management. Managers can have either full, partial or no discretion. We test the effect of these different forms of algorithmic-manager relationships on public employees' procedural justice. We hypothesize that the perceptions of procedural justice differ according to the extent to which issues are complex (Busch et al., 2018; Noordegraaf & Abma, 2003; Veale & Brass, 2019; Zarsky, 2016). We ask the following research question:

How does the inclusion of algorithms in managerial decision-making affect public employees' procedural justice perceptions of public management practices that differ in complexity?

The contribution of our work lies, first, in giving attention to using algorithms for managerial decisions in the public sector. Thus far, most attention has been directed at automating discretion at the frontline (Bovens & Zouridis, 2002; Busch & Henriksen, 2018; Reddick, 2005). Using algorithms for managerial decisions is an underexplored concept (Wesche & Sonderegger, 2019), even though key issues in the public sector, such as tension between rule following and discretion, are relevant at the managerial level as well (Maynard-Moody & Musheno, 2000). We also research the 'middle-ground', when algorithms serve as a decision-making partners rather than substitutes (Wesche & Sonderegger, 2019). This arrangement might be more realistic as, for instance, in Europe, Article 22 of the General Data Protection Regulation prohibits decision-making based on solely automatic processing (Finck, 2019). Through the inclusion of hybrid forms of decision-making, we extend the research by Lee (2018) on the effects of solely automating decisions in general management.

Second, we connect algorithmic public management to procedural justice. Algorithms can only be used for managerial decision-making if algorithms are perceived as legitimate (Wesche & Sonderegger, 2019). A lack of procedural justice can result in the rejection of using algorithms for certain management practices (Sunshine & Tyler, 2003). Thus, we believe that procedural justice has the potential to partly predict in which direction algorithmic public management will develop. In addition, procedural justice affects organizational variables relating to public employees' performance and well-being, such as job satisfaction, performance and organizational citizenship (Colquitt et al., 2001). As such, we explore the potential that including algorithms in managerial decisions has to make a positive or negative contribution. This connects to the societal responsibility of science to explore potential problems and opportunities in novel technological applications (Ghislieri et al., 2018). More generally, our paper contributes to the literature on algorithm aversion and antecedents of procedural justice within the public sector (Burton et al., 2019; Logg et al., 2019). We, moreover, contribute to research on public values as an important determinant of technology adoption, rather than just focusing on the technical aspects of technology (Lupo, 2019; Twizeyimana & Andersson, 2019a).

Third, we use an experimental approach. Experiments are especially valuable for detecting causal relationships (Gerber & Green, 2012; Margetts, 2011), they can account for unobserved confounders by randomization. Earlier research on perceptions within governmental organizations has used qualitative methods to detect complexity as an important factor in public employees' acceptance of discretion reduction (Busch et al., 2018). Our research contributes to testing this claim. We expand our experimental results by qualitatively assessing which aspects of complexity are most salient for public employees when algorithms are being used for managerial decisions.

The article will start by elaborating on algorithms, procedural justice, different types of algorithm-manager interactions and how perceptions are linked to management practices that differ in complexity. Then, we will present our hypotheses and explain our experimental method. Subsequently, we present our results and, finally, we end with a discussion and conclusion.

5.2 THEORY

5.2.1 ALGORITHMS

A technical definition of an algorithm is an 'abstract mathematical structure that has been implemented into a system for analysis of tasks in a particular analytic domain' according to Mittelstadt et al. (2016) adaption of Hill (2016, p.47). This definition consists out of two important elements. First, the algorithm refers to an *abstract mathematical structure*. Therefore, the algorithm does not imply necessarily the use of techniques, such as machine learning. It could be a simple linear model. Second, this structure has been configured into a *system for analysis*, such as a computer program or software.

The meaning of 'algorithms' yet goes beyond this technical definition. What an algorithm is, is constructed by the discourse surrounding algorithms and by the social context in which an algorithm is deployed (Beer, 2017). Burton et al., (2019) showed that, prior to contact with an algorithm, a human will have formed expectations about the algorithm. These expectations can be the product of experiences with algorithms, but also of reports from the media or peers. Research has shown that people can view algorithms as fair or unfair, irrespective of knowledge about the algorithm's procedure or accuracy (Lee, 2018). These perceptions are important as algorithms are often opaque, which causes us to lack information about the processes.

5.2.2. TYPES OF ALGORITHMIC PUBLIC MANAGEMENT

Algorithms and humans can interact in different ways (Jones, 2017; Rahwan, 2018). In this paper we use system-, screen- and street-level bureaucracy to describe different algorithm-manager interactions (Bovens & Zouridis, 2002; Zouridis et al., 2020). Bovens & Zouridis (2002) describe the changing role of public employees through the introduction of new technology. Broadly, the algorithm-manager interaction can take on three forms. First, algorithms can take over the role of the manager when algorithms are fully automated. In other words, technology is decisive. This leaves the manager with no discretion at all. There is little or no human judgment. This is called system-level bureaucracy. An example could be detecting potential problems in civil infrastructure (Spencer et al., 2019). The second scenario is when technology informs decisions, but a human decision maker is still required and able to exert judgment. We refer to this as screen-level bureaucracy. An example is predictive policing (Meijer & Wessels, 2019). The third scenario is the classic one in which technology is not necessarily used, but could serve as a support by choice. This is traditional street-level bureaucracy.

5.2.3. PROCEDURAL JUSTICE

More attention has been generated for the connection between justice and new technologies, for instance, in the smart city literature on e-justice (Lupo, 2019). Smart cities studies on e-justice have been criticized for mostly focusing on economic outcomes such as efficiency, while disregarding aspects relating to public value. Nevertheless, in the public sector, public values are essential for the successful adoption of technology (Twizeyimana & Andersson, 2019b). In this research, we focus on one of those public values - procedural justice (Page et al., 2015). Procedural justice recently has been connected to the task-technology fit in studies relating to digitalization in the public sector (Chen et al., 2016).

Procedural justice is part of a broader multi-dimensional justice construct consisting of distributive, procedural, informational, and interpersonal justice (Binns et al., 2018; Colquitt, 2001). These dimensions relate to different aspects of justice. Distributive justice, for example, covers one's assessment of the outcome of the decision. Procedural justice is specifically about the extent to which the process underlying decision-making is perceived as being fair (Colquitt, 2001; E. Lind & Tyler, 1988). Therefore, procedural justice does not necessarily correspond with one's assessment of the outcome. An outcome can be viewed as unfair, while the process with which the outcome was obtained is viewed as fair. Different aspects of justice can correlate (Binns et al., 2018).

Studying procedural justice is important for two main reasons. First, procedural justice has been connected to the use of algorithms because introducing algorithms might undermine the legitimacy of processes (Bovens & Zouridis, 2002; Citron & Pasquale, 2014; Crawford & Schultz, 2014; Danaher, 2016; Parkin, 2011). This can have different reasons as procedural justice is essentially an umbrella term that relates to perceptions on accuracy, consistency, bias suppression, correctability, representativeness and ethics (Greenberg & Colquitt, 2005). Therefore, a number of elements can contribute to procedural justice (Rubin, 2007), for instance, the extent to which an employee can voice opinions and can participate in decision-making. On top of that, the rules of the decision-making process, the process used to select those who make decisions and the existence of safeguards are also important (Leventhal, 1980). Second, it has been argued that procedural justice is essential for employees' positive job attitudes and behaviors in public sector organizations. Procedural justice can, moreover, have important effects on key organizational variables, including outcome satisfaction, job satisfaction, organizational commitment, trust, organizational citizenship behavior, withdrawal and performance (Colquitt et al. 2001).

5.2.4. COMPLEXITY IN PUBLIC MANAGEMENT

Theoretical models, such as the Technology Acceptance Model (TAM), the Unified Theory of Acceptance and Use of Technology (UTAUT) model and the Unified Model of Electronic Government Adoption (UMEGA), identify numerous factors that are important for technology adoption (Dwivedi et al., 2017; Venkatesh et al., 2003). These models focus on perceptions of technology's characteristics and the environment in which technology has to be implemented,

rather than the characteristics of the practice at hand. The task-technology fit model extends these technology adoption models by emphasizing the perceived fit between characteristics of the task and characteristics of the technology (Goodhue & Thompson, 1995).

There is a pervading sentiment in public administration that technology does not fit the nature of public administration, because decisions require human judgment (Oswald, 2018; Veale & Brass, 2019). To use algorithms for decision-making, we must be able to structurally measure, conduct and translate it into a model (Zarsky, 2016). Lipsky (2010) states that "the nature of service provision calls for human judgment that cannot be programmed and for which machines cannot substitute" (p.161). At the same time, rules have been at the center of public administration ever since Weber introduced his ideas on bureaucracy (Weber, 2015). As such, public administration is traditionally characterized by a tension between control and discretion (Busch et al., 2018; Evans & Hupe, 2019; Maynard-Moody & Musheno, 2000).

The call for human judgment in the public sector seems to be associated with the complexity of practices (Busch & Henriksen, 2018; Noordegraaf & Abma, 2003). There is no universally accepted definition of complexity (Mitchell, 2009). In this paper, we present two simplified dimensions of complexity: a technical and normative dimension. Complexity in a technical way boils down to practices consisting of many interconnected parts (Holland, 2014), which might be difficult to measure (Noordegraaf & Abma, 2003). Therefore, complexity can make a reductionist model of reality, in which more data might solve comprehensibility problems, difficult (Mitchell, 2009). Kallinikos (2005) proposed that the goal of contemporary technology is to make practices more manageable and predictable. Therefore, their success is connected to their ability to capture the processes for which they are designed (Kallinikos, 2009). However, complexity generally makes it more difficult to understand practices and predict outcomes. This relates ideas about how public services should be provided. Lipsky (2010), for example, explained that, for street-level bureaucrats, human judgment is necessary as most cases are unique. This uniqueness eliminates IF-THEN types of reasoning and requires discretion from bureaucrats.

Kallinikos (2009) extended his argument by stating that technology integration also depends on the context in which it needs to be embedded. This links to our second dimension of complexity. Complexity in public sector practices has a normative component (Noordegraaf & Abma, 2003). Practices can be contested. Comparable information will mean different things to different people. Trade-offs between values need to be taken into account and discretion should be used to deal with these trade-offs (Janssen & Kuk, 2016; Lipsky, 2010). This makes it impossible to objectively evaluate the 'right' course of action. As such, management cannot be optimal because what appears optimal differs from person to person. This makes practices multi-interpretable or equivocal.

We expect public employees to be in favor of more human judgment for managerial decisions as complexity increases (Busch et al., 2018). This relates to research of Lee (2018), who showed that for 'quantitative' tasks, such as work scheduling based on a

predicted amount of customers, algorithms and humans were perceived as equally fair. We expect that for low complexity tasks, which are conceptualized as quantitative tasks involving a limited number of variables and being relatively uncontested, public employees will be in favor of algorithmic management. In these cases, algorithms can increase accuracy and, using the most accurate option, is uncontested (Grove & Meehl, 1996). This leads to the following hypothesis:

Hypothesis 1

Decisions involving less human judgment are perceived higher in procedural justice for practices that are low in complexity.

For tasks involving emotions and human interaction, Lee (2018) showed humans were preferred over algorithms. In line with Lee, we expect that for highly complex practices, humans are preferred. This leads to hypothesis 2:

Hypothesis 2

Decisions involving less human judgment are perceived lower in procedural justice for practices that are high in complexity.

5.3. METHODS

We conducted two studies for this article. Our design builds on Lee's (2018) work, but extends it and specifies it to a public management context. Our groups represented three types of algorithmic-manager interactions based on Zouridis et al.'s (2019) typology of system-, screen- and street-level bureaucracy. We based our public management scenarios on real-life algorithmic management. Study 1 was pre-registered under <https://osf.io/xmzr8/>. In this between-subjects study, we researched perceptions of procedural justice in two different scenarios that varied in complexity. Study 2 was developed as a replication of the first study, but used a within-subject design. This allowed us to assess what effect multiple types of decision-making juxtaposed would have (Binns et al., 2018). This study also used different scenarios to test generalizability across various practices. All the conditions are shown in Appendix G. The experimental flow of the Study 1 can be found in Appendix H (Figure 9).

5.3.1. STUDY 1

In study 1 we studied two public management practices that are low and high in complexity. The practice low in complexity concerned determining how much reimbursement for commuting employees receive. In the Netherlands, rules concerning reimbursement for commuting public employees are settled in collective work agreements. Large organizations have additional rules for issues not covered in those agreements. As such, this practice follows a clear IF-THEN structure and relatively uncontested. This practice is thus low in complexity. Nevertheless, algorithmic advances are being made on travel cost

reimbursement, for instance, by employing carpool matching algorithms (Xia et al., 2015). We used performance evaluation of an employee for our high complexity scenario. Performance evaluation is a highly complex practice as it contains many variables of interest and the relevance of these variables for performance is contested (Dooren et al., 2015). Nevertheless, performance can be assessed by using algorithms, for example, in cases of teachers (Diakopoulos, 2014; O'Neill, 2016). We specified that these scenarios concerned back office employees of a municipality. They deal with requests on a case-to-case basis, which we see as central to public administration (Lipsky, 2010). We conducted a pilot study on Prolific. Our sample consisted out of 16 Dutch people, who did not necessarily work in the public sector. The procedure was not different than Study 1, with the exclusion of questions about demographics. We used a mixed ANOVA to estimate our power and expected a medium to large effect. We used the software program G*Power to conduct a power analysis. This provided a needed total sample size of 105.

5.3.2 STUDY 2

Study 2, the replication, used two different scenarios varying in complexity to establish generalizability. To reduce possible differences in interpretations of complexity, more information was provided about the factors contributing to the complexity of the practices. The low complexity scenario addressed the calculation of pensions in local government. These calculations are established through collective work agreements in the UK based on a limited amount of variables, such as inflation and pay (Kent Pension Fund, 2020). We considered this scenario as low in complexity as it has a limited number of variables and it is relatively uncontested. Simple calculation tools are available online.

The high complexity scenario involved the hiring of public employees. Selecting employees is a highly normative process that consists out of many intertwined factors (Villegas et al., 2019) and multiple phases (Uggerslev et al., 2012). Algorithms can be used in different phases of the hiring process, including CV screening and interviewing candidates through natural language processing and is contested (Binns et al., 2018; Raub, 2018).

In this study, we chose customer service officers as the type of employee. Our description of the job as a customer service officer was based on a real job ad. We used the PANGEA app for the calculation of our power based on a Cohen's d of 0,4 (Westfall et al., 2014). This led us to an estimation of 115 participants with 0.99 power. Scenarios were randomized to avoid learning/fatigue effects.

Other factors were kept constant. The process was opaque for both studies, practices possibly could have large consequences for employees, and outcomes were not specified. The conditions are shown in Appendix G.

5.3.3. MEASURES

Our dependent variable is procedural justice and consisted of a direct measure. Greenberg & Colquitt (2005) recommend using a direct measure when procedural justice is a dependent variable, and when event characteristics serve as the independent variable. We opted for a measure based on Lind & Tyler (1988) asking, for example, 'How fair is the procedure by which the performance of back office employees is evaluated?'. Respondents could then select a number from 1 ("Extremely unfair") to 7 ("Extremely fair") on a 7-point Likert scale. We then asked participants to explain their reasons for their ratings in an open-ended question. This allowed us to analyze to what extent complexity and elements of procedural justice played a role in the ratings for procedural justice. Our survey also measured subjects' managerial status (i.e., whether participants were in charge of managing subordinates), age, gender, educational background, and field of employment.

5.3.4. PROCEDURE

5.3.4.1. *Study 1*

Data was collected through an online survey using the alumni panel of the Utrecht School of Governance, which consisted mainly of public employees (75%). Participants were thus all higher educated and located across the Netherlands. We collected data for 109 Dutch public employees and excluded employees in the private sector from our analysis. Participation was voluntary. The survey was distributed through e-mail in the beginning of July 2019. Randomization was done in the Qualtrics survey software. We randomized participants to one of the three conditions. In this way, respondents received the two scenarios with the same algorithm-manager interaction. We made the decision-maker bold to emphasize how the decision was made. The manipulation check question asked was: "Which of the following made the decisions in the situations that you read?"

5.3.4.2. *Study 2*

Data was collected through an online survey using Qualtrics survey software and crowdsourcing platform Prolific Academic. Participants were pre-screened on residing in the UK and being an employee of a local, regional or national governmental organization. Each participant was presented with all scenarios and all possible decision-makers. We collected data on 126 public employees. The survey was distributed in the beginning of April 2020.

5.3.5. ANALYSES

For both studies, *p*-values are reported based on two-tailed hypotheses. We qualitatively analyzed participants' reasons for their answers to questions about procedural justice (Strauss & Corbin, 1990). We did this, first, by openly coding all quotes in our dataset. Afterwards, we axially coded our concepts by combining our codes into meaningful groups. The last step was to connect common codes to the concepts in our experimental design, leaving room for induction. The original data of Study 1 and Study 2 is available

online (<https://osf.io/xmzr8/>). The answers were translated into English and sentences were sometimes adapted to make them grammatically correct. Demographic data were removed from Study 1 to guarantee anonymity.

For Study 1, we conducted a mixed ANOVA. In case of an interaction effect, further ANOVAs on the separate conditions are conducted. Which conditions significantly differed, was analyzed by adopting the Games-Howell post-hoc test.

For Study 2, we conducted a two-way repeated measures ANOVA. When there was an interaction effect, we further proceeded to determine simple main effects per scenario by using repeated measure ANOVA's. Differences between conditions were tested using post-hoc *t*-test using the Bonferroni correction.

5.4. RESULTS

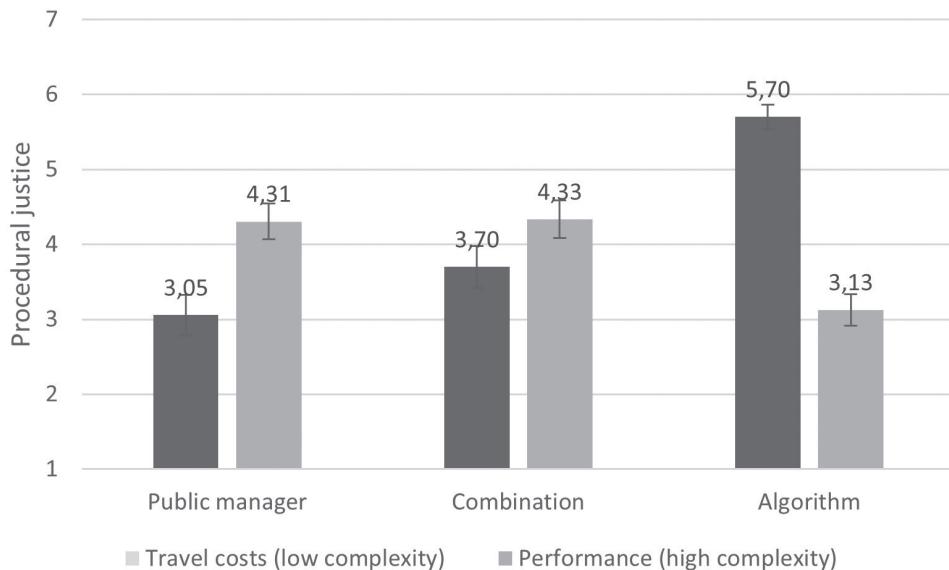
5.4.1. QUANTITATIVE ANALYSES

For study 1, the randomization check shows that our descriptive conditions are distributed equally among groups. All descriptives per group are shown in Appendix I (Table 13). The manipulation check indicates that the manipulation was successful ($\chi^2(4, N = 100) = 127.38 p = 0.00$). Exclusion of those who failed the manipulation check leads to similar results. Our mixed ANOVA presents a significant interaction effect of complexity and the different types of decision-making ($F(2, 106) = 44.09, p < 0.001$). Thus, the effect of how decisions are made on procedural justice is different in the two scenarios varying in complexity. We conducted two separate ANOVAs on both complexity conditions to interpret this effect. A significant omnibus result was found ($F(2, 107) = 35.21, p < 0.001$) for the scenario low in complexity on travel cost reimbursement. Post hoc comparisons using the Games-Howell test indicated significant differences between the algorithmic ($N = 40, M = 5.70, SD = 1.04$) and public manager condition ($N = 37, M = 3.05, SD = 1.65$) ($p < 0.001$, Cohen's $d = 1.96$), as well as for the algorithmic and combination condition ($N = 33, M = 3.70, SD = 1.61$) ($p < 0.001$, Cohen's $d = 1.53$). Differences between the combination and manager condition were not significant ($p = 0.23$, Cohen's $d = 0.40$). This indicates a medium effect size.

The ANOVA was significant for the scenario high in complexity as well ($F(2, 106) = 9.23, p < 0.001$). A post hoc Games-Howell test again indicated the algorithmic ($N = 40, M = 3.13, SD = 1.34$) and manager condition ($N = 36, M = 4.31, SD = 1.43$) to differ ($p < 0.001$, Cohen's $d = 0.86$), as well as the algorithmic and combination ($N = 33, M = 4.33, SD = 1.43$) condition ($p < 0.001$, Cohen's $d = 0.88$). The difference between the combination condition and the manager condition was not significant ($p = 1.00$, Cohen's $d = 0.01$). Results are displayed in Figure 7.

Figure 7

Means of procedural justice per group with standard error bars for study on a 7-point Likert scale.



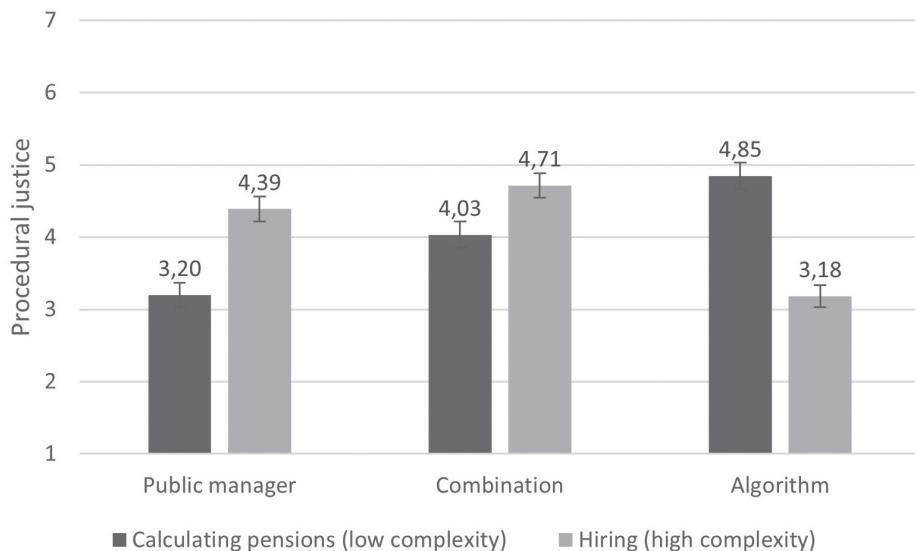
For Study 2, the within-subjects replication , the two-way repeated measures ANOVA indicated a significant interaction effect of complexity and conditions ($F(1.38, 172.45) = 103.68, p < 0.001$). Descriptives are shown in Appendix I (Table 14). Thus, the effect of the type of decision-making was again different in the two scenarios varying in complexity. We conducted a paired *t*-test to explore the main effect of complexity and two ANOVAs on both complexity conditions to interpret this effect.

For ANOVA on the scenario low in complexity, the calculation of pensions, showed a significant result ($F(1.4, 180.9) = 46.55, p < 0.001$) of the decision-maker on procedural justice. Post-hoc *t*-tests with the Bonferroni correction indicated significant differences between the algorithmic ($M = 4.84, SD = 2.08$) and public manager condition ($M = 3.20, SD = 1.90$) ($p < 0.001$, Cohen's $d = 0.74$), as well as the algorithmic and combination condition ($M = 4.03, SD = 2.02$) ($p < 0.001$, Cohen's $d = 0.38$) and the public manager and combination condition ($p < 0.001$, Cohen's $d = 0.70$).

For the scenario high in complexity concerning hiring, a significant result ($F(1.7, 218.5) = 59.65, p < 0.001$) was found as well of the decision-maker on procedural justice. Post-hoc *t*-tests with the Bonferroni correction indicated significant differences between the algorithmic ($M = 3.18, SD = 1.74$) and public manager condition ($M = 4.39, SD = 1.90$) ($p < 0.001$, Cohen's $d = 0.62$), as well as the algorithmic and combination condition ($M = 4.71, SD = 1.91$) ($p < 0.001$, Cohen's $d = 0.98$). The combination and manager condition were significantly different as well ($p = 0.035$, Cohen's $d = 0.22$). Results are displayed in Figure 8.

Figure 8

Means of procedural justice per group with standard error bars for study 2 on a 7-point Likert scale.



5.4.2. QUALITATIVE ANALYSIS

Our qualitative analyses revealed why these results occurred. We present the most important results below. First, we see that, in low complexity scenarios, algorithms are favored over managers, because managers are seen as subjective and biased, whereas computers using algorithms are seen as objective. To illustrate, respondents in Study 1 emphasized that judgment should be based on rules, such as the collective work agreement, and not on any one individual's judgment. In the algorithm condition, it was mentioned that the decision was based on "facts" or "hard data" and "rules", which are measurable. It was noted that the decision should be based on "established frameworks" (R73S1) and that "*It is not up to an individual within an organization to make such a decision*" (R24S1). Respondents in Study 2 mentioned that algorithms were favored over managers because "*managers are biased*" and computers are "*objective*" ensuring an equal process. R99S2 for instance noted "*Having a person making the decision means that he may be influenced by outside factors, whereas a computer would not be.*" And R40S2 noted "*The managers judgement may be unfair and not equal to others whereas a computer is not biased.*"

Nevertheless, some respondents mentioned it was positive that humans 'assisted' the algorithm, for instance, when algorithms err or when there are specific justifying circumstances. This explains why a combination is seen as more just in both studies for the low complexity decisions than decisions by a manager alone. R121S1, for example, said: "*If every employee is judged on the basis of equal criteria, it seems quite fair to me. If the personal situation requires a different treatment than that proposed by the computer, I think that the manager should be able to deviate from this.*" R115S2 mentioned "*Calculating*

pensions is a mathematically precise function with little room for human judgement. Using an algorithm for it is likely to produce more accurate and consistent results. There should always be room for human intervention to quality assure or check for errors, but the majority of the work should be calculated automatically." R15S2 mentioned "Judgement alone may be biased, and an algorithm alone may not take into account any special circumstances."

In high complexity conditions managers or manager-algorithm combinations are favored over solely using algorithms. This is mainly because soft aspects (such as compassion and fit within the team) and practices (such as face-to-face interviews in Study 2) are seen as being impossible for an algorithm to consider. R41S1 for instance mentioned "*A large part of the requests that come to them will be specific, and then compassion, solution-oriented approach, service, creativity, etc. are more important properties in proper functioning. That is difficult to measure with an algorithm.*" R76S2, for instance, mentioned "*Hiring relies far more on variables, which a computer algorithm is unable to process. Judging personality, values and fit with a team is not possible for algorithms.*" Some respondents mentioned that this required the skills from managers, which some respondents mentioned comes down to a "*feeling*".

Combinations of decision-makers are, in general, highly recommended in Study 1. Respondent R19S1 for instance mentioned: "*If everything functions like it should, the manager has a good view of the performance of its own employees. Then it is logical and good that he uses his own experience and judgment in the assessment. It would be good if that assessment is supplemented with more objective data and / or the opinion of others within the organization.*"

Nevertheless, combinations of algorithms and managers are only favored statistically in high complexity decisions when presented juxtaposed. In Study 2 respondents often mentioned that humans and computers have the ability to complement each other. An example is when R19S2 mentioned "*Humans and machines can make a mistake whilst working on their own. Working together ensures greater accuracy.*" However, in high complexity cases, the algorithm was more often preferred as assisting the human, instead of the other way around, as seen in the low complexity scenarios. Respondent R103S2 for instance mentioned "*A public manager can make a good judgement call, assisted by algorithm. A computer can't judge on a person's warmth and feeling.*"

5.5 DISCUSSION

Algorithms are increasingly applied in public management decision-making. However, this could be problematic for the legitimacy and acceptance of decisions. This paper sought to answer the question: *How does the inclusion of algorithms in managerial decision-making affect public employees' procedural justice perceptions of public management practices that differ in complexity?*

Our results have two main implications. First, public employees' procedural justice changes most when algorithms are fully automated and are replacing public managers. The direction of the effect, however, depends on the complexity of the practice at hand. For low complexity practices, automated algorithmic decision-making leads to higher reports of procedural justice, while the opposite effect occurs for high complexity practices. The attention for specific characteristics of the algorithm changes with the complexity of the practice. For low complexity practices, public employees indicate that they prefer algorithmic decision-making because it guarantees equal treatment across employees and is based on rules. The algorithms are seen as objective; thus, a controlled Weberian ideal of bureaucracy can be achieved by automating an algorithm (Weber, 2015). Decision-making by a public manager is less trusted, seen as subjective and reminds employees of special treatment. This differs from Lee (2018) who reported that for quantitative tasks, such as checking certain components of machinery, human and algorithmic decision-makers are viewed as equally fair. The quantitative tasks Lee (2018) researched can, however, be seen as more complex than the low complexity scenarios in this article. This indicates that the relationship between preferring an algorithm to a manager is linear based on complexity. In other words, less complexity of the practice means more preference for algorithms as a decision-makers. Apart from that, the difference between this article and Lee (2018) could be explained by this study's focus on procedural fairness, as opposed to general fairness, or could indicate that different values matter to public employees (Moore, 1995). Future research should replicate this study in the private sector.

For high complexity practices, decisions involving a public manager are perceived as higher in procedural justice than decisions that are made automatically by computers based on algorithms. This difference can be attributed the belief that algorithms cannot deal with practices that involve and attributed to rely on human skills, such as compassion and creativity. Public employees consider these factors as hard to measure. Thus, in highly technical complex practices, algorithms are not seen as being able to substitute for human judgment because of the nature of human interaction and service provision (Lee, 2018; Lipsky, 2010). Apart from that, public managers are trusted to judge performance and managers skills, which involve 'feeling', are welcomed in high complexity practices such as hiring and performance assessment. This corroborates Busch et al. (2018) findings that professional discretion is preferred for complex practices. As such, we expect algorithms to become at best decision-making partners, instead of algorithms becoming substitutes for public managers for highly complex practices (Wesche & Sonderegger, 2019).

Second, *adding* an algorithm to managerial decision-making can be beneficial. Benefits, nevertheless, depend on the practice at hand. For low complexity practices, hybrid decision-making is viewed more favorable than managers' judgment alone. For high complexity practices, a combination is only deemed more favorably when juxtaposed with the alternative of only having a public manager decide. The latter does not corroborate our second hypothesis and indicates the relationship between involving algorithms and procedural justice is not simply linear for highly complex practices. Future research

should test if these effects also occur in real-world scenarios, for instance, by testing the effect of adding an algorithm to hiring practices on public employees' procedural justice. Finally, we make two practical recommendations for public managers. The first recommendation is that automating decision-making in highly complex practices is not recommended, whereas automating low complexity practices is. When automating decision-making in high complexity practices is still preferred within organizations, public managers could pay attention to transparency and expectations in order to counter resistance to algorithmic decision-making. Transparency implies providing details on different aspects of an algorithm, such as platform design and algorithmic mechanisms (Ananny & Crawford, 2018). Transparency might be a solution, as transparency can generate trust (Kizilcec, 2016). However, a pitfall is that offering superficial information might give the illusion of transparency and not actually increase understanding of the workings of the algorithm (Janssen & Kuk, 2016). Apart from transparency, an open dialogue could help managers increase understanding of public employees' perceptions of algorithmic decision-making. In this way, public managers can take employees' expectations regarding decision-making into account.

Second, hybrid decision-making should be considered when public managers are operating alone. Moving from exclusively managerial to hybrid decision-making increases procedural justice for low complexity practices and, at least, does not decrease procedural justice for high complexity ones. In the latter case, the benefit of adding an algorithm will depend on different factors, such as trust in the manager and the performance of the algorithm. This might differ from case-to-case. More research on real life cases is needed to identify these factors and their relationships with each other.

5.6 LIMITATIONS

Our first limitations relate to the design. The scenarios varied on practice complexity. Complexity consists of a technical and a normative dimension. We treated complexity as one construct in this article. As such, our experimental research did not allow us to assess the causal effect of the individual dimensions. Future research should study these dimensions separately. In addition, some respondents indicated that, for them, a managers' judgment implied that the manager did not follow rules. Future research could test perceptions on managerial, hybrid and algorithmic decision-making when the same rules are explicitly followed. Lastly, our survey experiments suffered limitations similar to those reported in other online experiments (Manfreda et al., 2006).

Our second limitations relate to the scope of our research. This research only tests perceptions on the management of public employees, such as back office workers and customer service officers. Future research needs to test effects in other public sectors, such as healthcare, education and policing, or for other practices, such as predictive policing. Also, we focused on how algorithmic management was perceived when practices differed in complexity. Other factors can influence perceptions as well. Examples are the

number of cases that must be processed and if the organization is part of a centralized system (Bovens & Zouridis, 2002). Apart from that, future research should pay attention to the interplay between daily technology use at work and views on algorithmic decision-making (Scheerder et al., 2017).



CHAPTER 6

Conclusions and implications

6.1 INTRODUCTION

The rise of behavioral public administration demonstrated that we can understand and change decision-making by using insights in cognitive processes to study decision-making (Grimmelikhuijsen et al., 2017). Whereas earlier studies mainly referred to interventions, such as nudges, scholars have broadened their scope to include debiasing, and behavioral sciences beyond heuristics, such as the psychological COM-B model (Ewert, 2020). We are shifting our attention away from citizen-focused interventions to public sector worker-oriented interventions, i.e. the very people who are expected to nudge society (Hallsworth et al., 2018). This comes with unique challenges. Public sector workers operate within specific institutional settings and have to deal with particular situational settings (Noordegraaf, 2015). Not all public sector workers are, moreover, the same. For instance, a public manager might react differently to a behavioral intervention compared to a public employee. This dissertation seeks to explore how behavioral sciences can facilitate understanding and support decision-making across the public sector. We opted for a 'maximum variety' approach, including a range of behavioral theories, practices, issues and public sector workers to research the potential of using behavioral sciences in a broad way. In doing this, we allowed common themes to be identified throughout heterogeneity. We thus trace patterns related to 'handling heuristics' and we can highlight implications for subsequent research. We study the following overarching research question:

How can behavioral sciences facilitate the understanding and improvement of decision-making processes by public sector workers in public sector contexts?

This chapter first summarizes the answers to the sub-questions presented in chapters 2 to 5. Second, we answer the main research question by presenting three main conclusions. Third, the limitations of this dissertation are discussed. Fourth, the implications for academia and practice are assessed. Finally, we conclude on how to handle heuristics.

6.2 ANSWERING THE SUB-QUESTIONS

This section answers our sub-questions represented by the different chapters of the dissertation and the articles that were produced. We used a maximum variety approach to form the sub-questions and thus explored our main question from different angles. This allowed us to map a range of different behavioral theories, practices, public sector workers and issues. This maximum variety approach also offered possibilities to detect common themes despite of heterogeneity. The sub-questions were as follows:

1. How can nudges be used to support service provision, more specifically the provision of healthcare by healthcare professionals?
2. To what extent are public managers and employees affected by anchoring bias and how can this be reduced?

3. How can behaviors during political deliberations be understood, with an emphasis on active participation in online deliberations (during COVID-19)?
4. How can the procedural justice of management decisions be increased, especially when algorithms are used for decision-making?

Below, we answer these sub-questions.

6.2.1. NUDGES CAN SUPPORT SERVICE PROVISION DEPENDING ON CONTEXT (SUB-QUESTION 1)

Our first sub-question was '*How can nudges be used to support service provision, more specifically the provision of healthcare by healthcare professionals?*'. We studied this by conducting a systematic scoping review to map nudges used for evidence-based medicine. We found that nudging can influence service provision. In healthcare, some nudges, such as reminders, are often applied, while others, such as defaults, are rarely applied. This can be explained as reminders serve a purpose for the information overload that healthcare professionals experience (Allen & Harkins, 2005; Williamson et al., 1989). Nevertheless, other nudges, such as simplifying existing guidelines, can be studied more often (John & Blume, 2018; Michie & Lester, 2005). Nudges also target different barriers to behavior than information overload. Nudges might, for instance, focus on closing the intention-behavior-gap, on changing defaults in ordering systems or on the location of hand hygiene dispensers (Ansher et al., 2014; Kullgren et al., 2018). The outcomes that nudges targeted were mainly ordering, prescribing, and hand hygiene.

Nudging can thus support service provision. Yet, their success depends on institutional, as well as situational, settings. Institutional settings mattered, as the effects of interventions differ from organization to organization. Kiefe et al. (2001) for instance found that physicians working in rural areas are more reluctant to change decision-making based on performance feedback compared with suburban or urban physicians. Situational settings are important as nudging becomes more difficult as tasks become more complex. Nudges often target general outcomes, for example increasing healthcare professionals' hand hygiene. In these cases, nudging can be successful for relatively uncontested tasks. As tasks become more complex, general one-size-fits all nudges might no longer be applicable. Patel, Volpp, & Asch (2018), for instance, found that reducing the default duration of opioid prescriptions may make sense in acute conditions in an emergency department, but may be inappropriate for clinicians caring for patients with chronic pain. Apart from institutional and situational context, personal characteristics influence the effects of nudges. Less experienced doctors for instance reacted differently to nudges than more experienced doctors (Fogarty et al., 2013; Martens et al., 2007).

A way to contextualize and personalize nudges is to base them on algorithms. In healthcare, this happens by customizing nudges for specific clinical scenarios (see for instance Martens et al., 2007). An electronic patient file might for instance contain information about certain allergies of a patient, which can be used to tweak the nudge.

Evidence-based medicine has, however, been criticized for overly focusing on algorithmic rules that oversimplify clinical realities (Greenhalgh et al., 2014). Therefore, nudging might have its limits in healthcare and should always be done based on insights from healthcare professionals working with these clinical realities (Huis in 't Veld et al., 2020; Tummers, 2019b).

We conclude that nudges can support service provision of healthcare by healthcare professionals. Examples include: emphasizing the consequences for patients of proper hand hygiene; suggesting alternatives when clinicians propose antibiotics in digital environments; grouping tests or displaying them individually; changing the location of hand hygiene dispensers; and, putting reminders on operating room schedules. When applying nudging, one should always consider the specific institutional and situational settings. Nudges are not necessarily transferable from one organization to another, as institutional settings can influence a nudges' success. Nudges are also less suitable for complex medical issues. Target groups also need to be considered as some types of public professionals can react differently to nudges than others. Algorithms might offer solutions to contextualize and personalize nudges.

6.2.2. ANCHORING BIAS IS RELEVANT ACROSS CONTEXTS BUT REDUCING IT IS POSSIBLE (SUB-QUESTION 2)

Our second sub-question '*To what extent are public managers and employees affected by anchoring bias and how can this be reduced?*' was by conducting a replication survey experiment with public managers and employees. We extended this with a debiasing intervention. We found that decision biases, in this case anchoring, have an influence across different institutional contexts. Yet, institutional settings can affect how the bias influences decision-making. Our replication showed that our effect sizes differed from the original study. In our study effects were larger than in original studies. Apart from that, the mean minima and maxima shifted about 30 absolute points. This can be explained by multiple factors, including political reference levels, such as laws, having an effect on anchoring in the public sector (Holm, 2017).

We also found that the effect of anchoring bias can be reduced by using a debiasing intervention that is low in costs and in intensity. It is important that these interventions are adapted to the bias at hand. In our experiment, we used a consider-the-opposite strategy to debias decisions influences by anchoring (Mussweiler et al., 2000). This strategy achieved the desired outcome in all four cases. The effect sizes ranged from small to large, indicating that the effectiveness of the debiasing intervention is dependent on situational settings. In our case, the debiasing strategy was for instance more effective when a high anchor was presented. The reason for this could be that in this case the anchor was considered more extreme, which subsequently makes it easier to debias (Furnham & Boo, 2011).

The type of public sector worker also influences the effect of the consider-the-opposite intervention and seems to interact with the specific situational circumstances presented. Managers, for instance, adjust their anchors related to goal setting for their employees about responsiveness when they were asked to list reasons why the low anchor was inappropriate. They, however, did not adjust their answers when asked to do the same thing for a high anchor. On the other hand, employees did lower their anchors and moved to goals comparable with those of the managers. This could indicate that the decision-processes of employees and managers are different when confronted with a debiasing intervention. For example, managers might be more aware of the laws that are in place concerning responsiveness (Holm, 2017).

We conclude that anchoring bias is relevant across different institutional contexts, but its effect varies with these contexts. In other words, anchoring bias will not work the same everywhere and effects need to be considered by each case. Anchoring bias can be reduced by a low-cost and low-intensity intervention: consider-the-opposite. The effectiveness of consider-the-opposite can be different by each situational setting. The effect of the debiasing intervention might also depend on the public sector worker that is targeted. This means that the debiasing intervention might work differently on public managers compared with public employees.

6.2.3. COMPREHENSIVE BEHAVIORAL MODELS CAN HELP TO UNDERSTAND ACTIVE PARTICIPATION (SUB-QUESTION 3)

Our third sub-question '*How can behaviors during political deliberations be understood, with an emphasis on active participation in online deliberations (during COVID-19)?*' was studied through the case of online deliberation through videoconferencing by politicians in municipalities. We conducted semi-structured interviews and subsequently analyzed these interviews through a comprehensive behavioral model: the Theoretical Domains Framework (TDF) (Cane et al., 2012; Michie et al., 2005). The TDF is based on an aggregation of 33 psychological theories and consists out of 14 dimensions. In doing so, we researched which contextual settings and personal factors influenced their active participation within these meetings.

Context affects active participation. The influence of context becomes evident through institutional settings, for instance rules in meetings, but also situational settings, such as group polarization and group size. We found that, as reaction to the downsides of digital conferencing, rules were made that can create adverse choice architectures. Rules can for instance delay responding within a deliberation. This creates a barrier for active participation, partly because politicians believe responding is 'too little too late'. Situational settings also influence active participation. These situational settings can be group size and group polarization. When the group for instance becomes bigger, it is harder to maintain an overview of what is going on, which was already problematic in online video conferencing. When groups are polarized it can lead to more active participation, as people feel they have to contribute on things that matter to them.

Personal factors, such as skills with technology, power position and beliefs about if responding serves a purpose, affect active participation as well. Skills often are related to a person's age. Power position matters when, for example, a person without much power might think their opinion will not be taken into account. This means that some politicians, such as older politicians without much power, might participate less. This has implications for the representative and the democratic aspect of online deliberations (Brown, 2018; Moscrop & Warren, 2016).

We conclude that active participation during online deliberation can be understood through using a comprehensive behavioral model: the Theoretical Domains Framework (Cane et al., 2012; Michie et al., 2005). Context is influential through institutional settings, e.g. through rules in meetings, and through situational settings, such as group polarization and group size. In addition, personal characteristics such as skills with technology, a person's power position, as well as beliefs about if responding serves a purpose and goals affect active participation. Some of these factors offer possibilities for interventions, for example, if one can adjust rules or increase skills.

6.2.4. ALGORITHMS CAN BE USED FOR DECISION-MAKING, DEPENDING ON THE COMPLEXITY ISSUE AT HAND (SUB-QUESTION 4)

Our fourth sub-question was '*How can the procedural justice of management decisions be increased, especially when algorithms are used for decision-making?*'. We studied this by conducting two survey experiments to test how three types of algorithm-manager interactions influenced procedural justice perceptions in situations that were low and high in complexity. When algorithms are operating independently, procedural justice perceptions change the most. The direction of the effect is dependent on the complexity of the task. When a task is simple, procedural justice is higher when the decision is made by an algorithm as compared to managers. In other words, for relatively simple tasks algorithms are preferred over managers. When tasks are complex, procedural justice goes down and including managers for relatively complex tasks is preferred over solely using algorithms. This happens because for low complexity practices, public employees value equal treatment across employees based on rules. When practices are complex public employees value managerial discretion and do not believe that the algorithm can capture all the variables at play.

The results showed that adding an algorithm to managerial decision-making can increase procedural justice, but effects depend again on the situation at hand. When tasks are simple, *hybrid* decision making is favored over a manager deciding alone. When practices are complex, a combination is seen as more procedural just, but only when juxtaposed with only having a public manager decide. Otherwise, managerial and hybrid decision-making are seen as equally procedurally just. This means that the acceptability of changing decision-making processes should be viewed in relation to who, or what, is currently making the decisions.

The conclusion is that procedural justice can be changed by involving algorithms into managerial decision-making. The effect depends on the complexity of the task at hand, as well as previous forms decision-making. When a task is simple, algorithmic decision-making is preferred over managerial decision-making. When a task is complex, decision-making involving a public manager is preferred. Hybrid decision-making is preferred when hybrid decision-making is juxtaposed with only a manager making the decisions. Otherwise, hybrid and individual managerial decision-making are seen as equally procedurally just.

6.3 ANSWERING THE MAIN RESEARCH QUESTION

The following section will bundle these answers to answer our main research question. Our main research question was:

How can behavioral sciences facilitate the understanding and improvement of decision-making processes by public sector workers in public sector contexts?

We draw three main conclusions. First, behavioral sciences can support decision-making, most specifically by way of nudges, debiasing, conducting comprehensive analyses, and by using specific psychological theories. This can reap the rewards of insights on heuristics and biases by intervening in public sector decision-making. Behavioral sciences beyond heuristics can also be used. Second, when using these diverse behavioral tools, different institutional and situational settings in the public sector must be considered. Third, we need to personalize behavioral interventions by considering differences between public sector workers. We elaborate on these conclusions below.

6.3.1. BEHAVIORAL SCIENCES CAN SUPPORT DECISION-MAKING BY WAY OF NUDGES, DEBIASING, COMPREHENSIVE ANALYSES AND SPECIFIC PSYCHOLOGICAL THEORIES

The first main conclusion is that we can use behavioral sciences to support decision making in the public sector in four ways: by nudging, by adopting debiasing interventions, by using comprehensive behavioral analyses that inform interventions, and by using specific psychological theories. First, we can use *nudges* to support the decision-making of public sector workers. Choice architecture always has an influence on public sector workers' decision-making, even without anyone interfering in it (Vaughn & Linder, 2018). The choice architecture can however be consciously modified by installing nudges. Choice architectures are present in physical environments, and human interactions, and digital environments, such as online interfaces for ordering systems or e-mail systems. Many different applications of nudges are possible. Chapter 2 showed how healthcare professionals are nudged towards evidence-based medicine, including by providing reminders, by using social norm feedback about being a top performer, by displaying suggested alternatives, by pre-checking boxes, by organizing options differently, by

placing objects in the line of sights, by giving visual cues, by posters in hallways, and even through smells. These can be applied to public sector workers other than healthcare professionals. We can for instance imagine that the ease of accessing tools like documents will matter for frontline workers. We can also imagine that police officials will be sensitive to social norm feedback. Recent empirical evidence has corroborated this potential for application across different service contexts. For instance, Linos et al. (2017) changed information presented on post-cards to recruit new staff for police departments, and small financial incentives and reminders affected adoption of a program on providing information to the Ministry of Education in Denmark (Andersen & Hvidman, 2021). Likewise, Dustan et al. (2019) found that different text messages applying behavioral insights increased compliance with maintenance activities by civil servants.

We would also like to emphasize the potential for digital applications, in digital systems. Digital environments come with malleable choice architectures that will probably become more important in the future. This has to do with the digitalization of many public services. Chapter 5 described how technological developments are changing the way public services are provided. Interacting with digital systems is becoming increasingly important. In fact, as Zouridis et al. (2020) describe some public services are moving towards ‘screen’ and ‘system-level bureaucracy’, where public sector workers have less discretion, but rather interact with and respond to digital systems. An example is the rapid digital transformation can be observed in law, where online courts are increasingly used for decision-making (Donoghue, 2017). These online courts introduce digital choice architecture that might influence judgments (Sela, 2019). A pre-checked box might, for instance, result in a choice being selected more often, implying we will increasingly need to pay attention to how digital interfaces influence public sector decision-making. In doing this, we can draw from research on digital nudging or *e-nudging*, which is ‘the use of user-interface design elements to guide people’s behavior in digital choice environments’ (Weinmann et al., 2016, p.1).

Second, *debiasing* might be needed. Biases have an influence on decision-making by public sector workers (Battaglio et al., 2018). There are many examples of this (Hallsworth et al., 2018). For instance, framing bias does not so much concern the information presented, but rather how it is presented. This determines the course of decision-making. Chapter 3 illustrated that biases such as anchoring, which is the tendency estimate quantities by using an initial value (Tversky & Kahneman, 1974), can influence decision-making of public sector workers. This effect is robust across different institutional settings. This corroborates that we should take biases seriously when considering their influence on decision-making. Identifying how biases influence decision-making, especially in real world scenario’s, will therefore remain important.

Chapter 3 also showed that we can intervene upon biased decision-making by applying debiasing interventions that can be low-cost and low-in-intensity. When designing debiasing interventions we need to take into account that the mechanisms of the bias match the debiasing intervention (Croskerry et al., 2013a; Larrick, 2004). For instance, in

chapter 3 we saw that anchoring bias creates tunnel vision in which alternatives become inaccessible in the mind (Furnham & Boo, 2011). If we use consider-the-opposite, we break this tunnel vision (Mussweiler et al., 2000). Debiasing strategies like consider-the-opposite can also be applied to other biases which we know occur in public sector decision-making such as confirmation bias. For other biases, we might need other techniques (Hallsworth et al., 2018).

Third, *comprehensive analyses* can help understand issues from a behavioral perspective. Chapter 4 showed how to use the Theoretical Domains Framework (TDF) to analyze individual differences in active participation during online deliberations of council members (Michie et al., 2014; Michie et al., 2005). We found that personal as well as contextual factors were important in explaining active participation in online deliberations, and, subsequently, this can be used to create *tailored* interventions that could support decision-making. Chapter 4, for instance, described that goals were important. Goals could be targeted by an intervention such as goal setting or action planning (Michie et al., 2014). Other factors might require different interventions. These types of analyses are rarely used in public administration research. Yet, using them can give insights in what causes real world behavior in the public sector (Meder et al., 2018; Olejniczak et al., 2020; Weaver, 2015).

Fourth, we can continue and expand the use of *specific psychological theories* in public administration. Chapter 4 showed how to use procedural justice to research how the involvement of algorithms in managerial decisions changes perceptions for tasks differing in complexity. Thus, this theory is used to research attitudes regarding the increasing use of algorithms in the public sector. Many other individual psychological theories are applicable to public administration. Currently, scholars are expanding connections between behavioral sciences and public administration. For instance, recent work by Aleksovska (2021) has shown that accountability can be understood from a behavioral perspective. Other work has focused on citizens' stereotypes of street-level bureaucrats (de Boer, 2020).

This means that we can reap the rewards of insights on heuristics and biases by intervening in public sector decision-making by using nudges and debiasing interventions on public sector workers. We can also expand the use of behavioral sciences beyond heuristics. As such, the introduction of heuristics in public administration can be handled. Nudging, debiasing interventions, comprehensive models and specific psychological theories can be used depending on context. After conducting behavioral analyses, we might for instance conclude that we do not need to apply nudges and biases, but rather need to use other behavioral change techniques (Sousa Lourenço et al., 2016). Thus, we see heuristics and biases as part of the explanation of behavior, but not as a panacea. Different interventions can be used simultaneously as well. For instance, when implementing a new digital system, we might want to conduct a behavioral analysis to understand behavior, but at the same time consider the digital choice architecture.

6.3.2 BEHAVIORAL SCIENCES SHOULD BE CONTEXTUALIZED BY CONSIDERING INSTITUTIONAL AND SITUATIONAL SETTINGS

When working with behavioral interventions in the public sector, we must consider the institutional and situational settings in which behaviors occur. Chapter 2 illustrates that similar nudges can have different effects in different institutional settings. For example, a nudge might work in an urban practice, but not in a rural practice (Kiefe et al., 2001). A nudge can also be effective in the day shift but not in the night shift (Helder et al., 2012). We need more research to disentangle why this is the case. We do have indications that this happens because of the specific institutions relevant in those settings. Chapter 3 replicated an experiment in the UK that was previously executed in Italy showing that although the direction of the effect stayed the same, the effect sizes and the minima and maxima means differed substantively. This difference was bigger for our experiment concerning goal setting for responsiveness to citizens than for our experiment about performance appraisal. Chapter 3 explained how this can be caused by the institutions relevant to these two different practices. The differences between the UK and Italy can, for instance, be explained by different laws concerning the response time to citizens differ in the UK and Italy. As such, the effect of biases can be influenced by institutional settings.

Chapter 4 also showed that institutions can create choice architectures by the instalment of rules for behavior. More specifically, as a reaction to the downsides of digital deliberations, rules were made that created adverse choice architectures. Municipalities created rules about when and how to respond to each other. Politicians were for instance asked to respond after everyone had made his or her contribution. This caused people to experience barriers to respond: they were unable to remember all that had been said, and often felt that responding was ‘too little, too late’. As such, the rules set created psychological barriers that were difficult to overcome. This means that we need to study choice architecture beyond just tangible environments but also take the *institutions* of behavior into account. This might be especially relevant within organizations that inevitably consist out of institutions (March & Olsen, 2019).

Second, situational setting matters. We connected this to task complexity. The introduction and chapter 5, explain what complexity means. Complexity consists of two dimensions: a technical and a normative dimension. The technical dimension refers to a task consisting of many variables, that might be difficult to measure (Holland, 2014; Noordegraaf & Abma, 2003). The normative dimension refers to a task being contested (Noordegraaf & Abma, 2003). People then do not agree on what the outcome should be and how the task should be executed. Chapter 2 made it evident that nudges intervening on relatively uncontested, straightforward topics such as hand hygiene were the most effective. Nudges intervening on more complex topics were less effective. This might happen because nudges are often not contextualized, which makes them unfit for the complex realities of medical practice. This is corroborated by Huis in ’t Veld et al. (2020) who state that medical professionals think that the goals nudges should target must be as objective as possible. Using algorithms to contextualize nudges could be an option.

Chapter 5 researched what happens when algorithms are used for managerial decision-making and found that for practices low in complexity, algorithms are seen as appropriate for decision-making, while for practices high in complexity a manager or a hybrid form of decision-making is preferred. An algorithm might in those cases give a suggestion to a manager, which could be seen as a nudge (Vaccaro & Waldo, 2019). This implies that using nudges for high complexity cases is an option.

Chapter 5 also showed how complexity influenced cognitive processes more directly. When online deliberations became more complex because more actors were present, it was harder for people to maintain an overview. This caused politicians to be less aware of who said what in deliberations, and who they wanted to respond to. This means that complexity can amplify existing cognitive difficulties. Apart from that, the controversy of issues had an influence on if people wanted to actively participate. Politicians, for instance, were motivated to contribute when the topic was contested. This means that complexity has an influence beyond heuristics and biases.

This means, when handling heuristics and by extension behavioral sciences, that we have to consider institutional as well as situational settings. Concerning institutional settings, we have to explore the effects of institutions on the influence of nudges and biases. We can also consider how institutions create choice architectures. For situational settings, we can include the complexity of the task at hand.

6.3.3 BEHAVIORAL INTERVENTIONS CAN BE PERSONALIZED BY CONSIDERING INDIVIDUAL DIFFERENCES

The third main conclusion is that we can *personalize* behavioral interventions by considering individual differences in two ways. First, we can differentiate between sub-groups when applying nudges and debiasing interventions. Chapter 3 illustrated how public managers and employees reacted differently to our debiasing intervention, even to such an extent that while the intervention worked on public employees in one context, managers in the same context seemed unaffected by it. Employees for instance did not adjust their ratings after seeing a low-anchor combined with a consider-the-opposite intervention, while managers did. This implies that managers and employees have different ways of making decisions, and not all groups are affected equally by debiasing interventions. This corroborates work by for instance Christensen and Moynihan (2020) who found that debiasing politicians is harder than debiasing non-politicians.

Apart from distinguishing different types of occupations, we can distinguish between different types of professionals. Chapter 2 found that the type of professional can make a difference in how nudges are dealt with. Younger professionals might be more susceptible to nudges than older professionals. Recent research corroborates this by showing that professionals can perceive some nudges as interfering in their autonomy (Huis in 't Veld et al., 2020). This is related to their professionalism. In fact, the more professionalism, the stronger this effect. This indicates that professionalism and behavioral interventions, such as nudges, can be at odds with each other.

Second, we can tailor behavioral interventions to fit groups and individual's needs. This goes beyond adapting nudges and biases to different occupational and professional sub-groups. Rather, comprehensive behavioral analyses can identify psychological factors beyond heuristics and lead to different types of behavioral interventions. Chapter 4 described that individual differences can occur because of personal psychological factors such as skills, believing if responding serves a purpose and individual goals. These individual differences might vary per sub-group. Chapter 4 showed that concerning digital deliberations, older people often have fewer digital skills and might need more support. Algorithms might offer solutions. Algorithms are for instance being used to differentiate between the effectiveness of interventions on different sub-groups (Michie et al., 2017) (see for instance Richter et al., (2020) for an application in psychology). Algorithms are also applied in the private sector by companies to personalize interventions aimed at increasing productivity (Ebert & Freibichler, 2017).

These differences imply that nudging and debiasing as well as broader behavioral interventions are not a one-size-fits-all project, but rather would benefit from *personalizing* interventions. This can be done by targeting specific sub-groups by different nudges and debiasing interventions based on their decision-making styles. It can also be done by tailoring interventions based on their individual needs. Personalization opportunities based on algorithms might be a fruitful avenue.

6.4 LIMITATIONS

This dissertation has several limitations. First, we used different methods that have their own limitations. We have used survey experiments to test our hypotheses in chapter 3 and 4. Although these are useful for establishing causal claims, these experiments were conducted in an artificial setting, which could influence their external validity (Bouwman & Grimmelikhuijsen, 2016). Nevertheless, we used samples of real public managers and employees and based our scenarios on real world phenomena to ensure external validity. Additionally, our consider-the-opposite intervention and our requests for clarification both encompassed reflections, related to the real-world context. Future research will need to be applied in real world settings. Our qualitative study in chapter 5 on the other hand considers a real-world phenomenon based on the experiences of people working in a public sector environment. It identifies factors that can be intervened upon in a practical setting. Nevertheless, in this study, causality cannot be established, because it is based on semi-structured interviews. Future research should have an experimental character to test causality (Antonakis et al., 2010)

Second, although we use a maximum variety approach, we have not been able to cover the full range of topics pertinent to the public sector. We have also not been able to compare all different practices and issues to each other. Nevertheless, the aim of this dissertation was to identify common themes amidst heterogeneity. Future research should aim to compare different practices, issues and public sector workers, where this

is possible. We can, for instance, compare effects of nudges on different issues, or test nudges on different types of professionals.

Third, our main research question considers the influence of context, which is difficult to disentangle in specific research settings. This was evident in the experiments we reviewed in chapter 2. The systematic scoping review found that studies often did not consider different contexts or their influence. Replications of experiments, aimed at establishing generalizability across different context, have a similar problem. The replication in chapter 3 followed the method of the original experiments as closely as possible. Nevertheless, our experiments inevitably differed in time and space. This makes it hard to identify where the differences between our results exactly originate. Future scholars can include more context in experiments, by doing replications and sub-group analyses and by researching different issues or circumstances. The survey experiment of chapter 4, for instance, manipulated situational settings by manipulating low and high complexity of different tasks. We also can add different methods to get more in touch with different contexts. Evidence synthesis can be used to uncover conditions under with interventions are effective. Experiments can benefit from the addition of qualitative methods, and fully qualitative methods can be used to obtain rich in depth knowledge of how behavior takes place.

Context is very broad and can be defined in numerous ways (Pollitt, 2013). This can lead to problems because, as Wildavsky (1973) once stated about planning, if context is everything, maybe it is nothing. Context can, however, be specified, sometimes even into specific variables (Pollitt, 2013). In this study we focused on institutional and situational settings. The diversity of views on what context is, makes this inevitably limited (Pollitt, 2013). Yet, specifying context into concepts and variables contributes to making context more tangible for scholars and practitioners.

6.5 IMPLICATIONS

This section presents the implications of this dissertation for academia as well as practice. We first present five main implication for academia. These include future research suggestions for public administration scholars. Afterwards, we present four recommendations for practitioners.

6.5.1 ACADEMIC IMPLICATIONS

First, we recommend to distinguish between a narrow and broad definition of behavioral public administration (Ewert, 2020; Gofen et al., 2021). The *narrow* definition refers to using insights on heuristics and biases and, specifically, the diversion from rational models of decision-making. The *broad* definition considers cognitive processes in general that can influence decision-making. These different definitions can be beneficial to specify which streams of literature we are drawing upon. This differentiation is also useful to broaden our scope into cognitive processes beyond heuristics. We provide two examples of how this broadened scope can be applied. Chapter 4 explores the topic of

active participation during videoconferencing by applying a multifaceted comprehensive model of psychological factors to the data. This identified a diverse set of factors that influenced active participation. Similar research can also apply this, or other models, to analyze determinants of behavior (Olejniczak et al., 2020). Researchers can also apply specific psychological theories. We used procedural justice as a specific psychological theory. Many more applications are possible, including those which have already been applied, such as public service motivation (Vandenabeele, 2007). Yet, some theories about cognitive processes remain still under-researched. This includes connecting cognitive processes to neurological and physiological responses (for examples, see Han et al., 2021; Hattke et al., 2020)).

Second, we have to consider more precisely how we can use behavioral sciences specifically in public sector contexts (Ewert et al., 2021; Feitsma & Whitehead, 2019). This relates to the possible schism that was previously warned for in public administration research (Moynihan, 2018). Having a ‘schism’ implies a binary view of micro and macro approaches to public administration. However, as in the quantitative/qualitative debate, this binary approach can be argued to be not very beneficial (Gilad, 2019). Instead, most research will cover different levels, without even realizing it. However, research needs explicitly to include context, either in data collection, or in the discussion to expose these important connections. Albeit context has a wide range of possible definitions, this needs to be done by at least specifying *which aspects of context* matter for the research at hand and *how* this context makes a difference (Pollitt, 2013).

This dissertation emphasizes that institutional and situational settings are influencing decision-making. Institutional settings, in the form of institutions, such as laws and more informal rules, can influence the effects of heuristics and biases on decision-making. Institutions itself can also shape choice architectures. Situational settings are influential as well. We find that the complexity of a task has implications for the effects of heuristics and biases. Complexity can also influence cognitive processes and behaviors through other mechanisms, such as motivation. This dissertation uses a maximum variety approach to identify common themes in a broad range of topics. Future research will need to specifically include different institutions and task complexities in their methodologies to explore their effects in combination with behavioral factors.

Third, we need to differentiate between various public sector workers, given their diversity in terms of occupations, professionalism and other personal factors. These differences can have implications for interventions effectiveness and require us to move beyond one-size-fits all approaches when intervening on public sector workers. Rather, we can *differentiate* at the minimum between occupational groups. We can, for instance, include professionalism, which is currently under-researched, when studying behavioral sciences in public sector contexts. In addition, decision-making is determined by other personal psychological factors, such as skills, beliefs about consequences, goals, and motivations. These factors can be used to personalize interventions, for instance, by catering to different people’s needs. This can be done by combining behavioral interventions, such as nudges, with algorithms (Mills, 2020).

Fourth, we can use different methodologies to explore behavioral factors influencing decision-making in public administration. Current approaches to behavioral public administration mostly consist of survey experiments that are used to explore the influence of behavioral factors on decision-making within the public sector (Battaglio et al., 2018; Bhanot & Linos, 2020). We find that although survey experiments can add new knowledge, other methodologies can be used as well, as they have unique advantages. For example, the systematic review bundles existing evidence and shows how nudges are being applied within public organizations. The replication contributes to knowledge about generalizability across different institutional contexts. The survey experiments combined with qualitative data gave insight into the mechanisms behind the quantitative data. The qualitative study, with semi-structured interviews allowed us to make a comprehensive behavioral analysis of a novel phenomenon within the public sector. We thus encourage the use of methods beyond survey experiments.

Fifth, behavioral sciences can be more explicitly linked to digitalization. Digitalization is becoming increasingly relevant, as we move towards system and screen-level bureaucracy (Zouridis et al., 2020). Scholars are, moreover, emphasizing errors in human decision-making and possible solutions to those can be found in digital systems (Kahneman et al., 2021). Digitalization and public administration can be linked through the use of algorithms in several ways. Algorithms can be used to contextualize and personalize interventions. Contextualization is possible by adapting interventions to available contextual data, such as patient information. We can also use algorithms to personalize nudges and debiasing interventions for specific occupational and professional groups or individuals. Additionally, algorithmic suggestions have been labelled as nudges themselves as they steer behavior but do not dictate it (Vaccaro & Waldo, 2019). As such, we have only just begun to explore the power of combining algorithms with behavioral interventions related to heuristics and biases (Yeung, 2017). Finally, behavioral sciences can help to understand innovations. Algorithmic public management can, for instance, be understood through studying effects on procedural justice and individual differences in dealing with digital innovations, such as online deliberation, can be analyzed by applying comprehensive behavioral frameworks.

6.5.2 SOCIETAL IMPLICATIONS

The main conclusions can be translated into at least four recommendations for practitioners. Our first recommendation is to look for opportunities to apply nudges within public organizations to support decision-making of public sector workers. Chapter 2 explained how this happens within health care organizations. Nudging can happen, for example, through providing reminders, through comparing health care professionals to each other, through changing defaults for test ordering, through asking for accountable justifications and through putting medical tools in the line of sight. These can be applied to different public services as well. Recent research, for example, experiments with letting judges know that they are going to be reviewed by a third person or that they will have to explain their decision-making process (Schillemans & Giesen, 2020), reminding school staff of the deadline for opting in or out for a national program (Andersen & Hvidman,

2021) and providing a diagram emphasizing vulnerabilities during the life-course for frontline workers judging disability insurance (Visintin et al., 2021).

The second recommendation is to be aware of possible sources of biased decision-making and apply debiasing interventions. Biases can be present in almost all stages of the policy process. They can affect how people interpret information, how people interact within groups, and also how policies are implemented (see Battaglio et al., 2018; Hallsworth et al., 2018; Moseley & Thomann, 2021 for more examples of biases in public-decision making). For designing a debiasing intervention, it is essential to understand the mechanism by which the bias works. This dissertation focused on a debiasing intervention to counter anchoring. We found that a consider-the-opposite intervention was effective as this removes the anchoring bias induced 'tunnel vision' (chapter 3). Consider-the-opposite possibly also could be applied to other narrow thinking biases such as confirmation bias (Soll et al., 2015). Other biases might need different debiasing techniques such non-face to face input to counter group reinforcement (Hallsworth et al., 2018), or reframing techniques, by presenting both the negatively and positively framed information for framing bias (Cantarelli et al., 2018).

The third recommendation is to apply other psychological models to understand decision-making. Using a multifaceted general model connected to specific behavioral interventions such as the Theoretical Domains Framework or the COM-B model can be a good starting point (Michie et al., 2014). Semi-structured interviews can uncover the mechanisms that lead to the (non) display of behaviors that can be connected to interventions. Behavioral analyses might suggest nudging public sector workers, but can also suggest other behavioral interventions (Sousa Lourenço et al., 2016). Practitioners can also use specific psychological theories to explain decision-making. This dissertation used procedural justice to understand perceptions on algorithmic public management. Practitioners can also draw on other theories, from psychology in general, to specific psychological theories within public administration such as public service motivation, policy alienation, and red tape (see Bozeman, 1993; Tummers, 2011; Vandenabeele, 2007).

The fourth recommendations is to take institutional and situational settings into account, as well as personal characteristics of public sector workers. When applying interventions one can – for instance – ask which rules are applicable to the decision-making and how they affect the intervention. These rules can be formal and written, or informal and social. When applying interventions, the situational setting needs to be considered. It may be that nudges are, for instance, not positively viewed because of the complexity of the situation. Nudges based on algorithms could provide solutions to this, but they should be used with caution. Public sector workers might not view the substitution of human decision makers by algorithms for complex decisions as procedurally just. Also, different public sector workers might have different needs and/or might respond to interventions differently. These differences can be based on occupations, on professionalism but also on other individual differences, such as skills. This calls for always including the people who are going to be intervened upon in the process (Huis in 't Veld et al., 2020; Tummers, 2019a).

6.6 CONCLUDING HOW TO HANDLE HEURISTICS

This dissertation has researched how public administration researchers and practitioners can handle the introduction of heuristics in public administration to support decision-making of public sector workers. It focused on exploring the meaning of behavioral sciences and proposed a narrow and broad definition. It also explored how public sector context can be included in behavioral public administration. Therefore, the main question we asked was: *How can behavioral sciences facilitate the understanding and improvement of decision-making processes by public sector workers in public sector contexts?* We addressed this by adopting a maximum variety approach to research different practices, issues, public sector workers and behavioral theories. This allowed us to map a broad range of applications of behavioral sciences, while still having the opportunity to identify common themes amidst heterogeneity.

This dissertation found that nudges, debiasing, comprehensive behavioral models and specific psychological models can be used to understand and support decision-making processes by public sector workers. Understanding heuristics and biases can lead to interventions that support public sector decision-making. Yet, we also have to consider other psychological factors beyond heuristics and biases affect individual decision-making.

It is important that behavioral public administration is contextualized and personalized. This dissertation illustrated that including context can be done by concentrating on the institutional and situational settings public sector workers operate in. More specifically, we can include institutions, such as rules, and task complexity in research designs, descriptions of context and subsequently in implementation decisions.

Apart from that, individual differences between public sector workers cannot be ignored. Different public sector workers can have different needs and can react differently to behavioral interventions. We can personalize interventions by differentiating between occupational and professional sub-groups. Behavioral analyses can help us identify specific personal factors that influence behavior. Algorithms might help us contextualize and personalize behavioral public administration. We can, for instance, use algorithms to offer interventions based on the specific case at hand in digital systems, and we can personalize interventions based on a public sector workers' specific needs. Using algorithms for decision-making about complex issues should, however, be done with a public manager. Further research is needed to explore the potential of using algorithms combined with behavioral sciences. As such, we can handle heuristics to understand and support decision-making in public sector contexts.

SUMMARY

SUMMARY

INTRODUCTION

What if we could support public sector decision-making through low-cost, low-intensity interventions? This is the promise of 'nudging' public sector workers. A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2008, p. 6). An example in public policy is changing the default of organ donation from being opt in to opt out (Jachimowicz et al., 2019; Johnson & Goldstein, 2003). Nudges are based on heuristics and biases. Heuristics are mental shortcuts that reduce complex tasks to simpler ones (Kahneman et al., 1982; Tversky & Kahneman, 1974). Sometimes heuristics are very useful, for instance, when people have to make fast decisions in a crisis situation. They can, however, also create bias, which means that decision-making becomes suboptimal. An example is when public sector workers' decision-making is influenced by irrelevant numbers (Bellé et al., 2018). In these situations, interventions can be used to mitigate these effects. These are called debiasing interventions (Larrick, 2004).

Behavioral sciences are sometimes equated with the use of heuristics and biases (Sousa Lourenço et al., 2016). Yet, behavioral sciences can be argued to encompass more than just heuristics and biases. This relates to a narrow and broad view of behavioral sciences (Ewert, 2020; Gofen et al., 2021). A narrow view entails focusing on heuristics and biases. A broad definition of behavioral sciences is where cognitive processes lead to decision-making. When we use the broad definition, our scope of research increases drastically to include studies beyond heuristics and biases from, for instance, psychology. Using psychology to study the behavior of public sector workers is not new. In healthcare, scholars have used psychology to study the implementation of evidence for many years. This evidence has recently been bundled into comprehensive models such as the Theoretical Domains Framework. This bundles 33 psychological theories into 14 factors including intentions, skills and emotions (Cane et al., 2012; Michie et al., 2005).

In public administration, psychology has been used before as well. In the 1940s Herbert Simon already emphasized the importance of including psychology in the study of public administration (Simon, 1997, p. xi). Although this explicit integration did not take off, we can observe the inclusion of theories related to heuristics, biases, and psychology. An example is the work of Stone (1997) on the framing of policy stories. However, behavioral public administration, which describes combining psychology with public administration, was not coined until 2016 (Grimmelikhuijsen et al., 2017). This happened partly because integrating psychology and public administration obtained a new impulse after the widespread academic and governmental attention for behavioral interventions such as nudges. The behavioral public administration stream is now maturing. Scholars are now for instance focusing on behavioral theories beyond heuristics and biases (Ewert et al., 2021; Gofen et al., 2021). Scholars are also moving beyond just studying citizens and are turning their attention to studying the decision-making of public sector workers instead (Hallsworth et al., 2018).

Focusing on public sector workers instead of citizens raises questions about how the public sector context is important for the application of behavioral sciences. Unlike citizens, public sector workers often operate within organizations and groups (Jones, 2017). In this dissertation, we focus on two aspects of context: institutional settings and situational settings. Institutional settings can be researched by concentrating on institutions, i.e. rules that shape decision-making (March & Olsen, 1983). Situational settings are important when tasks differ in complexity (Noordegraaf & Abma, 2003; Noordegraaf, 2015).

Complexity can occur because tasks consist of a large number of factors. Imagine the construction of a single street versus the construction of a whole neighborhood. However, tasks can also be complex because they differ in being contested. Some topics might be relatively uncontested, such as maintaining roads. Other topics might be very contested, for example, raising taxes. Apart from these institutional and situational settings, differences between public sector workers are important to consider. These differences can be based on occupation, but also on other factors such as skills and motivations.

Therefore this dissertation answers the following main research question:

How can behavioral sciences facilitate the understanding and improvement of decision-making processes by public sector workers in public sector contexts?

RESEARCH APPROACH

This dissertation is based on four articles. This dissertation uses a maximum variety approach to answer the main research question. This means that the main question is approached from various angles. It covers different behavioral theories, practices, issues, and public sector workers. This allows us to explore the full range of applications of behavioral sciences to public sector workers decision-making. With this approach common themes can be detected that receive their validity as a result of being present amidst heterogeneity. In Table 8, we present an overview of our chapters, methodologies, the sub-questions and the status of the articles.

Table 8

An overview of the chapters of this dissertation, their methodology, corresponding sub-questions and their status

Chapter	Methodology	Sub-question	Status
2: Nudging healthcare professionals towards evidence based medicine: A systematic scoping review	Systematic scoping review of 100 articles	How can nudges be used to support service provision, more specifically the provision of healthcare by healthcare professionals?	Published in <i>Journal of Behavioral Public Administration</i>
3: Designing to Debias: Measuring and Reducing Public Managers' Anchoring Bias	Survey experiment: Replication and extension with 1221 public managers and employees	To what extent are public managers and employees affected by anchoring bias and how can this be reduced?	Published in <i>Public Administration Review</i>
4: Don't speak: exploring how behavioral factors affect politicians' participation in online deliberations	Semi-structured interviews with 8 politicians and 7 council clerks	How can behaviors during political deliberations be understood, with an emphasis on active participation in online deliberations (during COVID-19)?	Under review
5: The impact of using algorithms for managerial decisions on public employees' procedural justice	Two survey experiments combined with qualitative data with 235 public sector workers	How can the procedural justice of management decisions be increased, especially when algorithms are used for decision-making?	Published in <i>Government Information Quarterly</i>

THE CONTRIBUTION

This dissertation contributes to academia and practice. For academia first, we emphasize a narrow *and* broad view of behavioral sciences, and by extension, of behavioral public administration. In doing this we shed light on the potential of using behavioral sciences beyond heuristics and biases. Second, we move beyond the focus on citizens, and join recent efforts join recent efforts focused on decision-making inside public organizations (Bellé et al., 2018; Hallsworth et al., 2018; Soman & Yeung, 2021). While doing this we include the public sector context in our studies. The importance of context has been emphasized in public administration in general as well as specifically for behavioral public administration (Jilke et al., 2019; Moynihan, 2018; Pollitt, 2013; Roberts, 2020). Context is broad and not easy to define (Pollitt, 2013). This dissertation concentrates on institutional and situational settings. We also differentiate between individual public sector workers.

Third, methodologically, we show that different methods are useful to study how behavioral sciences can be used in the public sector. This diverges from the earlier focus of behavioral public administration on strengthening experimental methodology within public administration research (Bhanot & Linos, 2020; Grimmelikhuijsen et al., 2017). We do this by using a systematic scoping review to synthesize evidence, a replication to establish generalization, a mixed methods study to causality test and detect underlying mechanisms, and a qualitative study with semi-structured interviews to explore a novel topic. As such we show how scholars can use different methods beyond just survey experiments in behavioral public administration.

This dissertation also contributes to practice in various ways. This dissertation for instance offers practical tools for practitioners to apply behavioral sciences and subsequently change decision-making within the public sector. These tools are nudges to apply on public service workers as well as a debiasing intervention. We also show how comprehensive behavioral analysis can be used to identify personal and contextual factors that influence decision-making.

Apart from all this, the issues we focus on hold their own specific relevance for society. We study evidence based medicine, which refers to healthcare professionals making decisions about patients using the best available evidence (Greenhalgh et al., 2014). This approach to medicine can lead to more cost-effectiveness and better health care (Masic et al., 2008). Evidence-based medicine has been regarded as an utopia because of information overload (Allen & Harkins, 2005; Williamson et al., 1989). Nudging, possibly, can offer solutions to this. We also study how anchoring bias affects goal setting and performance appraisal. Goal setting and performance feedback are two types of core internal management practices in the public sector (Favero et al., 2016; Pedersen & Stritch, 2018a). We also study how we can debias anchoring effects. As such, we not only identify problems but also attempt to solve them. This dissertation also studies the inclusion of different politicians in deliberations by studying active participation of politicians. This contributes to knowledge about how we can safeguard representative democracy within online deliberative democracy, specifically through videoconferencing by politicians (Albrecht, 2006; Brown, 2018; Friess & Eilders, 2015). Because of the COVID-19 crisis, public decision-making bodies had to move to online spaces. The digital divide literature emphasizes inequalities between individuals that can exist as a result of these spaces (Scheerder et al., 2017; Van Deursen & Van Dijk, 2019). By undertaking a comprehensive behavioral analysis, we identify factors contributing to active participation to harbor equality in online deliberations (Michie et al., 2014). Finally, we study the potential of algorithms to take over or be involved in managerial tasks relating to human resource management within the public sector. Algorithms are increasingly used for managerial decision-making (Wesche & Sonderegger, 2019). Yet, we do not know how public employees view their effect on procedural justice. This study also sheds light on the potential of algorithms to contextualize and personalize behavioral interventions as well as solving problems of 'noise' (Kahneman et al., 2021; Mills, 2020).

S

ANSWERS TO THE SUB-QUESTIONS

We now present the answers to the sub-questions.

Nudges can support service provision depending on context (Sub-question 1)

Nudging (i.e. softly steering) healthcare professionals towards utilizing evidence-based medicine may be a feasible possibility, but the potential of using nudges depends on context. We studied this by conducting a systematic scoping review to map nudges used towards evidence-based medicine. We found that nudging can influence service

Summary

provision. In healthcare, some nudges, such as reminders, are often applied, while others such as defaults are rarely applied. This can be explained by reminders serving a purpose, given the information overload that healthcare professionals experience (Allen & Harkins, 2005; Williamson et al., 1989). Nevertheless, other nudges, such as making existing guidelines easier by simplifying them, can be researched more often (John & Blume, 2018; Michie & Lester, 2005).

Thus, nudging can support service provision, but effectiveness depends on institutional, as well as situational settings. Situational settings are important as we see that nudging becomes more difficult as tasks become more complex. Nudges often target general outcomes; an example is increasing healthcare professionals' hand hygiene. In such cases, nudging can be successful for relatively uncontested tasks. Apart from institutional and situational context, personal characteristics influence the effects of nudges. For example, less experienced doctors reacted differently to nudges compared to more experienced doctors (Fogarty et al., 2013; Martens et al., 2007).

A way to contextualize and personalize nudges is to base them on algorithms. In healthcare, this happens by customizing nudges based on specific clinical scenarios (see for instance Martens et al., 2007). An example could be where an electronic patient file might contain information about certain allergies of a patient, which can be used to tweak the nudge. However, evidence-based medicine has been criticized for overly focusing on algorithmic rules that oversimplify clinical realities (Greenhalgh et al., 2014). Therefore, nudging might have its limits in healthcare and should always be done based on insights from healthcare professionals working with these clinical realities (Huis in 't Veld et al., 2020; Tummers, 2019b).

Anchoring bias is relevant across contexts but reducing it is possible (Sub-question 2)

Anchoring bias has effects across institutional context, but reducing it with a debiasing intervention is possible. We studied this by conducting a replication through a survey experiment on public managers and employees. We extended this with a debiasing intervention. This dissertation found that public managers' decisions are affected by anchoring bias. For instance, employees' previous year's performance ratings influence new ratings irrespective of actual performance. The results indicate that anchoring bias replicates in a different institutional context, although effect sizes differ. Furthermore, a low-cost, low-intensity consider-the-opposite technique mitigates anchoring bias in this survey experiment. An exploratory sub-group analysis indicates that the effect of the intervention depends on institutional settings, such as the laws that are relevant for the task at hand. The type of public sector worker confronted with the bias or debiasing intervention also seems to matter. A next step is to test the consider-the-opposite strategy in real-world settings.

Comprehensive behavioral models can help to understand active participation

(Sub-question 3)

We can use comprehensive behavioral models to understand behavior. In chapter 4, this was applied to study active participation of politicians during online deliberation. We conducted semi-structured interviews and subsequently analyzed these interviews through a comprehensive behavioral model: the Theoretical Domains Framework (TDF) (Cane et al., 2012; Michie et al., 2005). The COVID-19 crisis caused public decision-making bodies to deliberate online. Online environments are known to lead to inequalities because of differences not only access, but also in the way people actively participate online. We, moreover, know that personal and contextual behavioral factors might be able to explain these differences, while simultaneously offering possibilities for intervention. Our results show that personal factors (for instance beliefs about if responding serves a purpose), and contextual factors (such as rules in meetings) influence active participation. We recommend practitioners to offer organizational support to account for a lack of skills, to intervene upon beliefs about consequences and to make fitting rules by sharing best practices.

Algorithms can be used for decision-making, depending on the complexity issue at hand

(Sub-question 4)

Algorithms can be used for managerial decision-making in the public sector, but possibilities for application depend on the complexity of the issue at hand. We studied this by conducting two survey experiments in which we tested how three types of algorithm-manager interactions had an influence on procedural justice perceptions in situations that were low and high in complexity. We found that when algorithms are operating independently, procedural justice perceptions change the most. The direction of the effect is dependent on the complexity of the task. When a task is relatively simple, procedural justice is higher when the decision is made by the algorithm as compared to managers. In other words, algorithms are preferred over managers. When tasks, however, are complex, procedural justice goes down. In other words, managers are preferred over solely using algorithms. This happens because for low complexity practices, public employees value equal treatment across employees based on rules. When practices are complex public employees value managerial discretion and do not believe that the algorithm can capture all variables at play.

We also found that adding an algorithm to managerial decision-making can increase procedural justice, but effects depend again on the situation at hand. When tasks are simple, *hybrid* decision making is favored over a manager deciding alone. When practices are complex, a combination seen as more procedural just when combinations are juxtaposed with only having a public manager decide. Otherwise, a manager and hybrid decision-making are seen as equally procedurally just. This means that the acceptability of changing decision-making processes should be viewed in relation to who, or what, is currently making the decisions.

THREE MAIN CONCLUSIONS

On the basis of the answers to our sub-questions we draw three main conclusions.

Behavioral sciences can support decision-making by way of nudges, debiasing, comprehensive analyses and specific psychological theories

Our first main conclusion is that we can use behavioral sciences to support decision making in the public sector in four ways: by nudging, by adopting debiasing interventions, by using comprehensive behavioral analyses that inform interventions, and by using specific psychological theories. First, we can use *nudges* to support the decision-making of public sector workers. Many different applications of nudges are possible. In chapter 2 we saw that healthcare professionals are nudged towards evidence-based medicine. This was done, for instance, by providing reminders, by using social norm feedback about being a top performer, and by displaying suggested alternatives. This can be applied to public sector workers other than healthcare professionals. Recent evidence has highlighted the effect of small financial incentives and reminders on the adoption of a program on providing information to the Ministry of Education in Denmark (Andersen & Hvidman, 2021). We would also like to emphasize the potential for digital applications, through digital systems. Digital environments come with malleable choice architectures that will probably become more important in the future. This has to do with the digitalization of many public services (Bovens & Zouridis, 2002). In doing this we can draw from research on digital nudging or *e-nudging* which is ‘the use of user-interface design elements to guide people’s behavior in digital choice environments’ (Weinmann et al., 2016, p.1).

Second, *biases* might have an influence on decision-making by public sector workers (Battaglio et al., 2018). In chapter 3 we saw that biases such as anchoring, which is the tendency to estimate quantities by using an initial value (Tversky & Kahneman, 1974), can influence decision-making of public sector workers. This effect is robust across different institutional contexts. In chapter 3, we also saw that we can intervene upon biases decision-making by applying a low-cost and low-intensity debiasing intervention. When designing debiasing interventions we need to take into account that the mechanisms of the bias match the debiasing intervention (Croskerry et al., 2013a; Larrick, 2004). For instance, in chapter 3 we saw that anchoring bias creates tunnel vision in which alternatives become inaccessible in the mind (Furnham & Boo, 2011). If we use consider-the-opposite, we break this tunnel vision (Mussweiler et al., 2000). Debiasing strategies can also be applied to other biases we know occur in public sector decision-making such as confirmation bias. For other biases, we might need other techniques (Hallsworth et al., 2018).

Third, *comprehensive analyses* can help understand issues from a behavioral perspective. In chapter 4, we used the Theoretical Domains Framework (TDF) to analyze individual differences in active participation during online deliberations of council members (Michie et al., 2014; Michie et al., 2005). We found that personal, as well as contextual, factors were important in explaining active participation in online deliberations. This analysis can be

used subsequently to create *tailored* interventions that could support decision-making. In chapter 4, we for instance found that goals were important. Goals could be targeted by an intervention such as goal setting or action planning (Michie et al., 2014). Other factors might require different interventions. These types of analyses are rarely used in public administration research, yet using them can give insights in what causes real world behavior in the public sector (Meder et al., 2018; Olejniczak et al., 2020; Weaver, 2015).

Fourth, we can continue and expand the use of *specific psychological theories* in public administration. In chapter 5, we use procedural justice to research how the involvement of algorithms in managerial decisions changes perceptions for tasks differing in complexity. Scholars have recognized this potential and have been researching different aspects of behavioral sciences in public administration. Notable examples include research on red tape, on public service motivation, and on policy alienation. Currently, scholars are expanding connections between behavioral sciences and public administration. An example is recent work by Aleksovska (2021) that has shown that accountability can be understood from a behavioral perspective.

Behavioral sciences can be contextualized by considering institutional and situational settings

When working with behavioral sciences in the public sector, we have to consider the institutional and situational settings in which behaviors occur. First, the institutional setting can make a difference. We have indications that this happens because of the specific institutions. In chapter 3, we for instance replicated an experiment in the UK that had previously been executed in Italy. We found that although the direction of the effect stayed the same, the effect sizes and the minima and maxima means differed substantively. This difference was bigger for our experiment concerning goal setting for responsiveness to citizens than for our experiment about performance appraisal. In chapter 3 we explained that this can be caused by the institutions relevant to these two different practices. The differences between the UK and Italy can for instance be explained by different laws concerning the response time to citizens differ in the UK and Italy. As such, the effect of biases can be influenced by institutional settings.

In chapter 4, we also saw that institutions can create choice architectures, by installing rules for behavior. More specifically, as a reaction to the downsides of digital deliberations, rules are made that created adverse choice architectures. Municipalities created rules about when and how to respond to each other. Politicians, for instance, were asked to respond after everyone had made his or her contribution. However, this caused people to experience barriers to respond: they were not able to remember all that had been said, and often felt like responding was ‘too little, too late’. As such, the rules set create psychological barriers that are difficult to overcome. This means that we need to study choice architecture beyond just tangible environments but also take the *institutions* of behavior into account. This might be especially relevant within organizations that inevitably consist out of a plethora of institutions (March & Olsen, 2019).

Second, we showed that the situational setting matters. This was mostly related to the complexity of the task. In chapter 2 it became evident that nudges intervening on relatively uncontested, straightforward topics, such as hand hygiene, were the most effective. Nudges intervening on more complex topics were less effective. This might happen because nudges are often not contextualized, which makes them unfit for the complex realities of medical practice. This is corroborated by Huis in 't Veld et al. (2020) who state that medical professionals have the opinion that the goals that nudges target should be as objective as possible. Using algorithms to contextualize nudges could be an option. In chapter 5, we researched what happens when algorithms are used for managerial decision-making. We found that for practices low in complexity, algorithms are seen as appropriate for decision-making, while for practices high in complexity a manager or a hybrid form of decision-making is preferred. An algorithm, in those cases, might give a suggestion to a manager, which could be seen as a nudge (Vaccaro & Waldo, 2019). This implies that using nudges for high complexity cases is an option.

In chapter 5, we also saw that complexity influenced cognitive processes more directly. When online deliberations became more complex because more actors were present, it was harder for people to keep an overview. This caused politicians to be less aware of who said what in deliberations, and who they wanted to respond to. This means that complexity amplifies existing cognitive difficulties. This means that complexity has an influence beyond heuristics and biases.

Behavioral interventions can be personalized by considering individual differences

Our third main conclusion is that we can *personalize* behavioral interventions by considering individual differences. We can do this in two ways. First, we can differentiate between different sub-groups when applying nudges and debiasing interventions. In chapter 3, we saw that public managers and employees reacted differently to our debiasing intervention. This implies that managers and employees have a different way of making decisions, and not all groups are affected equally by debiasing interventions. Apart from distinguishing different types of occupations, we can distinguish between different types of professionals. In chapter 2, we found that the type of professional might make a difference for how nudges are dealt with. For example, younger professionals might be more susceptible to nudges than older professionals. This indicates that professionalism and behavioral interventions, such as nudges, can be at odds with each other.

Second, we can tailor behavioral interventions to fit groups and individual's needs. This goes beyond adapting nudges and biases to different occupational and professional sub-groups. Rather, comprehensive behavioral analyses can identify psychological factors beyond heuristics and lead to different types of behavioral interventions. In chapter 4, we showed that individual differences can occur because of personal psychological factors such as skills, believing if responding serves a purpose and the goals that a person has. These individual differences might vary by sub-group. In chapter 4, we saw that for concerning digital deliberations, older people often have less digital skills and might need more support. Algorithms might offer solutions for both personalization options.

This can be done by differentiation by sub-groups but also by personalizing interventions (Ebert & Freibichler, 2017; Michie et al., 2017).

IMPLICATIONS

These three main conclusions have implications for academia as well as practice. First, for academia, we recommend to distinguish between a narrow and broad definition of behavioral public administration (Ewert, 2020; Gofen et al., 2021). The *narrow* definition refers to using insights on heuristics and biases and specifically the diversion from rational models of decision-making. The *broad* definition considers cognitive processes in general that can influence decision-making. Second, we have to consider more precisely how we can use behavioral sciences specifically in public sector contexts (Ewert et al., 2021; Feitsma & Whitehead, 2019; Jilke et al., 2019). This dissertation showed how specifically institutional and situational settings can be added to behavioral public administration research. Third, we need to differentiate between various public sector workers. Public sector workers are diverse in terms of occupations, professionalism and other personal factors. Because of this, we need to move beyond merely one-size-fits-all approaches to intervening on public sector workers. Rather, we can differentiate and personalize interventions. Fourth, we can use different methodologies to explore behavioral factors influencing decision-making in public administration beyond the survey experiment. Rather we can use systematic (scoping) reviews, replications and extensions, mixed methods and qualitative methods. Fifth, behavioral sciences can be more explicitly linked to novel topics relating to digitalization. These topics are very relevant at the moment, as we move towards system and screen-level bureaucracy (Zouridis et al., 2020). Algorithms can moreover be understood through behavioral sciences as well as offer heterogeneity in interventions based on contextual and individual differences.

For practice, the first recommendation is to look for opportunities to apply nudges within public organizations to support decision-making of public sector workers. This can happen by modifying physical environments such as office buildings, on physical objects such as forms, human interactions, but also in digital environments. Our second recommendation is to be aware of possible sources of biased decision-making and apply debiasing interventions. These biases can be present in almost all stages of the policy process. They can be present in how people interpret information, how people interact within groups, but also in how policies get implemented (see Battaglio et al., 2018; Hallsworth et al., 2018; Moseley & Thomann, 2021 for more examples of biases in public-decision making). For designing a debiasing intervention, it is essential to understand the mechanism by which the bias works. In our work, the consider-the-opposite intervention was for instance effective to debias anchoring, but other biases might need different debiasing interventions. Our third recommendation is to apply other psychological models to identify novel decision-making situations or issues in existing decision-making situations. Using a multifaceted general model, which can be connected to specific behavioral interventions such as the Theoretical Domains Framework or the COM-B model can be a good starting point (Michie et al., 2014). Our fourth recommendation is

Summary

to ensure institutional and situational settings are taken into account, as well as personal characteristics of public sector workers.

CONCLUSION

We conclude that scholars and practitioners can handle the introduction of heuristics by conceptually acknowledging a narrow and broad definition of behavioral sciences. Empirically, we can study and use nudges, debiasing interventions, comprehensive behavioral models and specific psychological theories. In doing this, we need to include institutional and contextual settings. We also need to differentiate between public sector workers and personalize interventions. Algorithms might help contextualize and personalize behavioral public administration. Further research is needed to explore the potential of using algorithms combined with behavioral sciences in public administration. As such, we can handle heuristics to understand and support decision-making in public sector contexts.

S

SAMENVATTING IN HET NEDERLANDS

SAMENVATTING IN HET NEDERLANDS

INTRODUCTIE

Wat als we besluitvorming in de publieke sector zouden kunnen ondersteunen door middel van goedkope interventies met een lage intensiteit? Dit is de belofte van het 'nudgen' van werknemers in de publieke sector. Een nudge is "elk aspect van de keuzearchitectuur dat het gedrag van mensen op een voorspelbare manier verandert zonder opties te verbieden of hun economische prikkels aanzienlijk te veranderen" (Thaler & Sunstein, 2008, p. 6). Een voorbeeld in overheidsbeleid is het veranderen van de standaard optie van orgaandonatie van 'nee' naar 'ja' (Jachimowicz et al., 2019; Johnson & Goldstein, 2003). Nudges zijn gebaseerd op heuristieken en biases. Heuristieken zijn mentale snelkoppelingen die complexe taken terugbrengen tot eenvoudigere taken (Kahneman et al., 1982; Tversky & Kahneman, 1974). Soms zijn heuristieken erg handig, bijvoorbeeld wanneer mensen snel beslissingen moeten nemen in een crisissituatie. Ze kunnen echter ook bias creëren, waardoor de besluitvorming suboptimaal wordt. Een voorbeeld is wanneer de besluitvorming van werknemers in de publieke sector wordt beïnvloed door irrelevante cijfers (Bellé et al., 2018). In deze situaties kunnen interventies worden ingezet om deze effecten te verminderen. Dit worden 'debiasing' interventies genoemd (Larrick, 2004).

Gedragswetenschappen worden soms gelijkgesteld met het gebruik van heuristieken en biases (Sousa Lourenço et al., 2016). Gedragswetenschappen omvatten echter volgens sommige academici meer dan dat. We kunnen een beperkte en brede kijk op gedragswetenschappen onderscheiden (Ewert, 2020; Gofen et al., 2021). Een beperkte kijk op gedragswetenschappen houdt in dat we ons concentreren op heuristieken en biases. Een brede definitie van gedragswetenschappen omvat cognitieve processen die leiden tot besluitvorming. Wanneer we de brede definitie hanteren, neemt ons onderzoeksgebied drastisch toe. Dan kunnen we onderzoek uitvoeren dat verder gaat dan heuristieken en bias. We kunnen bijvoorbeeld theorieën uit de psychologie gebruiken om het gedrag van werknemers in de publieke sector te bestuderen. Dit is niet nieuw. In de gezondheidszorg gebruiken wetenschappers bijvoorbeeld al jarenlang psychologie om de implementatie van bewijs te bestuderen. Dit bewijs is onlangs gebundeld in uitgebreide modellen zoals het Theoretical Domains Framework dat 33 psychologische theorieën bundelt in 14 dimensies, waaronder intenties, vaardigheden en emoties (Cane et al., 2012; Michie et al., 2005).

In de jaren veertig benadrukte Herbert Simon ook al het belang van het opnemen van psychologie in de bestuurskunde (Simon, 1997, p. xi). Hoewel deze expliciete integratie niet van de grond kwam, zijn theorieën met betrekking tot heuristieken, biases en psychologie al eerder gebruikt in de bestuurskunde. Een voorbeeld is het werk van Stone (1997) over het framen van beleidsverhalen. De gedragsbestuurskunde, dat de combinatie van psychologie met openbaar bestuur beschrijft, werd echter pas in 2016 geïntroduceerd (Grimmelikhuijsen et al., 2016). Dit kwam mede doordat de integratie van psychologie en bestuurskunde een nieuwe impuls kreeg na de wijdverbreide academische en

bestuurlijke aandacht voor gedragsinterventies zoals nudges. Gedragsbestuurskunde is nu volwassen aan het worden. Wetenschappers richten zich op dit moment bijvoorbeeld op gedragstheorieën die verder gaan dan heuristieken en biases (Ewert et al., 2021; Gofen et al., 2021). Wetenschappers gaan ook verder dan alleen het bestuderen van burgers en richten hun aandacht in plaats daarvan op het bestuderen van de besluitvorming van werknemers in de publieke sector (Hallsworth et al., 2018).

De focus op werknemers in de publieke sector, in plaats van op burgers, roept vragen op over hoe de context van de publieke sector belangrijk is voor de toepassing van gedragswetenschappen. In tegenstelling tot burgers opereren werknemers in de publieke sector vaak binnen organisaties en groepen (Jones, 2017). In dit proefschrift richten we ons op twee aspecten van context: de institutionele en situationele omgeving. We kunnen het effect van de institutionele omgeving onderzoeken door ons te concentreren op instituties, d.w.z. regels die de besluitvorming vormgeven (March & Olsen, 1983). De situationele omgeving is belangrijk wanneer taken verschillen in complexiteit (Noordegraaf & Abma 2003; Noordegraaf 2015).

Complexiteit kan ontstaan doordat taken uit een groot aantal factoren bestaan. Stel je bijvoorbeeld de aanleg van een enkele straat voor versus de aanleg van een hele wijk. Taken kunnen echter ook complex zijn omdat ze controversieel zijn. Sommige onderwerpen kunnen relatief onomstreden zijn, zoals het onderhoud van wegen. Andere onderwerpen kunnen echter zeer omstreden zijn, zoals het verhogen van belastingen. Afgezien van deze institutionele en situationele omgeving, zijn verschillen tussen werknemers in de publieke sector belangrijk. Deze verschillen kunnen betrekking hebben op het beroep dat mensen uitvoeren, maar ook op andere persoonlijke factoren zoals vaardigheden en motivaties.

Daarom beantwoordt dit proefschrift de volgende hoofdonderzoeksvraag:

Hoe kunnen gedragswetenschappen het begrip en de verbetering van besluitvormingsprocessen door werknemers in de publieke sector in publieke sector contexten bevorderen?

ONDERZOEKSAANPAK

Dit proefschrift is gebaseerd op vier artikelen. In tabel 9 presenteren we een overzicht van vier verschillende hoofdstukken die betrekking hebben op de verschillende artikelen. Deze dissertatie gebruikt een maximale variëteitsbenadering om de hoofdonderzoeksvraag te beantwoorden. Dit betekent dat de hoofdvraag vanuit verschillende invalshoeken wordt beschouwd, waarbij verschillende gedragstheorieën, praktijken, problemen en werknemers in de publieke sector worden behandeld. Dit stelt ons in staat om het volledige scala aan toepassingen van gedragswetenschappen voor de besluitvorming van werknemers in de publieke sector te verkennen. Toch kunnen met deze benadering ook gemeenschappelijke thema's worden opgespoord die hun belang ontlenen aan hun aanwezigheid te midden van heterogeniteit. In tabel 9 geven we een overzicht van onze hoofdstukken, methodieken, deelvragen en de status van de artikelen.

CONTRIBUTIE

Dit proefschrift heeft een meerwaarde voor de wetenschap en de praktijk. Voor de academische wereld leggen we, ten eerste, de nadruk op een beperkte en brede kijk op gedragstheorieën, en in het verlengde daarvan, op gedragsbestuurskunde. Zo werpen we licht op het potentieel van het gebruik van gedragstheorieën, voorbij alleen heuristieken en bias. Ten tweede gaan we verder dan de focus op burgers, door ons aan te sluiten bij recente inspanningen gericht op besluitvorming binnen publieke organisaties (Bellé et al., 2018; Hallsworth et al., 2018; Soman & Yeung, 2021). Daarbij betrekken we de context van de publieke sector in ons onderzoek. Het belang van context is benadrukt in bestuurskundig onderzoek in het algemeen en specifiek in voor de gedragsbestuurskunde (Jilke et al., 2019; Moynihan, 2018; Pollitt, 2013; Roberts, 2020). Het begrip context is breed en niet gemakkelijk te definiëren (Pollitt, 2013). Deze dissertatie concentreert zich op de institutionele en situationele omgeving. We maken ook onderscheid tussen individuele werknemers in de publieke sector.

Tabel 9

Een overzicht van de hoofdstukken van dit proefschrift, toegepaste methodologie, bijbehorende deelvragen en hun status

Hoofdstuk	Methodologie	Deelvragen	Status
2: Nudging healthcare professionals towards evidence based medicine: A systematic scoping review	Systematic scoping review van 100 artikelen	Hoe kunnen nudges worden ingezet ter ondersteuning van dienstverlening, in het specifiek voor de zorgverlening door zorgprofessionals?	Gepubliceerd in <i>Journal of Behavioral Public Administration</i>
3: Designing to Debias: Measuring and Reducing Public Managers' Anchoring Bias	Survey experiment: Replicatie en uitbreiding met 1221 publiek managers en werknemers	In hoeverre hebben publieke managers en medewerkers last van anchoring bias en hoe kan dit worden verminderd?	Gepubliceerd in <i>Public Administration Review</i>
4: Don't speak: exploring how behavioral factors affect politicians' participation in online deliberations	Semi-gestructureerde interviews met 8 politici en 7 griffiers	Hoe kunnen gedragingen tijdens politieke beraadslagingen worden begrepen, met de nadruk op actieve deelname aan online beraadslagingen (tijdens COVID-19)?	Under review
5: The impact of using algorithms for managerial decisions on public employees' procedural justice	Twee survey experimenten gecombineerd met kwalitatieve data onder 235 publiek werknemers	Hoe kan de procedurele rechtvaardigheid van managementbeslissingen worden vergroot, in het specifiek wanneer algoritmen worden gebruikt voor besluitvorming?	Gepubliceerd in <i>Government Information Quarterly</i>

Deze dissertatie draagt ook op verschillende manieren bij aan de praktijk. Dit onderzoek biedt bijvoorbeeld praktische handvatten om gedragstheorieën toe te passen en daarbij de besluitvorming binnen de publieke sector te veranderen. Deze instrumenten zijn nudges toe te passen op ambtenaren die diensten verlenen, evenals een 'debiasing'

interventie. We laten ook zien hoe uitgebreide gedragsanalyse kan worden gebruikt om persoonlijke en contextuele factoren te identificeren die de besluitvorming beïnvloeden.

Daarnaast hebben devraagstukken waar we ons op richt een specifieke maatschappelijke relevantie. We bestuderen evidence-based medicine, wat verwijst naar professionals in de gezondheidszorg die beslissingen nemen over patiënten met behulp van het best beschikbare bewijs (Greenhalgh et al., 2014). Deze benadering van geneeskunde kan leiden tot betere gezondheidszorg en meer kosteneffectiviteit (Masic et al., 2008). Evidence-based medicine wordt door de hoeveelheid aan informatie als een utopie beschouwd (Allen & Harkins, 2005; Williamson et al., 1989). Nudging kan hier mogelijk oplossingen voor bieden. We bestuderen ook hoe 'anchoring bias' het stellen van doelen en prestatiebeoordeling beïnvloedt. Het stellen van doelen en prestatiebeoordeling zijn twee typen kernpraktijken voor intern management in de publieke sector (Favero et al., 2016; Pedersen & Stritch, 2018a). We bestuderen ook hoe we 'anchoring bias' kunnen verminderen. Zo signaleren we niet alleen problemen, maar proberen we ze ook op te lossen. Deze dissertatie bestudeert daarnaast de inclusie van verschillende politici in online deliberatie. Dit draagt bij aan kennis over hoe we representatieve democratie binnen online deliberatieve kunnen waarborgen, met name tijdens videoconferenties door politici (Albrecht, 2006; Brown, 2018; Friess & Eilders, 2015). Vanwege de COVID-19-crisis moesten openbare besluitvormende organen verplaatsen naar online ruimtes. Literatuur over digitale kloof benadrukt de ongelijkheden tussen individuen die kunnen bestaan door gebruik van deze online ruimtes (Scheerder et al., 2017; Van Deursen & Van Dijk, 2019). Door een uitgebreide gedragsanalyse uit te voeren, identificeren we factoren die bijdragen aan actieve deelname om gelijkheid in online beraadslagingen te bevorderen (Michie et al., 2014). Ten slotte bestuderen we het potentieel van algoritmen om managementtaken, met betrekking tot human resource management, binnen de publieke sector over te nemen of erbij betrokken te worden. Algoritmen worden steeds vaker gebruikt voor bestuurlijke besluitvorming (Wesche & Sonderegger, 2019). We weten echter niet hoe ambtenaren hun effect op procedurele rechtvaardigheid zien. Deze studie werpt bovendien ook licht op het potentieel van algoritmen in het algemeen om gedragsinterventies te contextualiseren en te personaliseren en om problemen van 'ruis' op te lossen (Kahneman et al., 2021; Mills, 2020).

ANTWOORDEN OP DE DEELVRAGEN

We presenteren nu de antwoorden op de deelvragen.

Nudges kunnen dienstverlening ondersteunen afhankelijk van de context (deelvraag 1)

Nudging (zacht sturen) van zorgprofessionals om evidence-based medicine te bevorden kan een haalbare mogelijkheid zijn, maar het potentieel van nudges is afhankelijk van de context. We hebben dit onderzocht door via een systematische scoping review nudges voor evidence-based medicine in kaart te brengen. We ontdekten dat nudging de dienstverlening kan beïnvloeden. In de gezondheidszorg worden sommige nudges, zoals geheugensteuntjes, vaak toegepast, terwijl andere, zoals standaardinstellingen, zelden

worden toegepast. Dit kan worden verklaard door dat geheugensteuntjes de informatie-overload die zorgprofessionals ervaren kunnen verminderen (Allen & Harkins, 2005; Williamson et al., 1989). Toch kunnen andere nudges, zoals het vereenvoudigen van bestaande richtlijnen, vaker worden onderzocht (John & Blume, 2018; Michie & Lester, 2005).

Nudging kan dus de dienstverlening ondersteunen, maar de effectiviteit hangt af van zowel de institutionele als situationele omgeving. De situationele omgeving is belangrijk omdat nudging moeilijker wordt naarmate taken complexer worden. Nudges zijn vaak gericht op algemene uitkomsten: een voorbeeld is het stimuleren van het regelmatig handenwassen van zorgprofessionals. In die gevallen kan nudging succesvol zijn. Naast de institutionele en situationele omgeving, beïnvloeden persoonlijke kenmerken de effecten van nudges. Minder ervaren artsen reageerden bijvoorbeeld anders op nudges dan meer ervaren artsen (Fogarty et al., 2013; Martens et al., 2007).

Een manier om nudges te contextualiseren en te personaliseren, is door ze te baseren op algoritmen. In de zorg gebeurt dit door nudges aan te passen op basis van specifieke klinische scenario's (zie bijvoorbeeld Martens et al., 2007). Een elektronisch patiëntendossier kan bijvoorbeeld informatie bevatten over bepaalde allergieën van een patiënt, waarmee de nudge kan worden aangepast. Evidence-based medicine is echter bekritiseerd omdat het te veel focust op algoritmische regels die de klinische realiteit te eenvoudig maken (Greenhalgh et al., 2014). Daarom heeft nudging in de gezondheidszorg grenzen en moet het altijd worden gedaan op basis van inzichten van zorgprofessionals die met deze klinische realiteiten werken (Huis in 't Veld et al., 2020; Tummers, 2019b).

Anchoring bias is relevant in verschillende contexten, maar verminderen van bias is mogelijk (deelvraag 2)

Anchoring bias heeft effecten in verschillende institutionele omgevingen. Het is echter mogelijk om anchoring bias te verminderen met een debiasing interventie. We hebben dit onderzocht door een replicatie uit te voeren via een survey-experiment onder publieke managers en werknemers. We hebben dit uitgebreid met een debiasing interventie. Deze dissertatie laat zien dat de beslissingen van publieke managers worden beïnvloed door anchoring bias. Zo beïnvloeden de prestatiebeoordelingen van werknemers van vorig jaar nieuwe beoordelingen, ongeacht de werkelijke prestaties. De resultaten geven aan dat anchoring bias replicateert in een andere institutionele omgeving, alhoewel de effectgroottes verschillen. Bovendien vermindert een goedkope, makkelijke, 'consider-the-opposite' techniek, de anchoring bias in dit survey-experiment. Een verkennende subgroepanalyse geeft aan dat het effect van de interventie afhangt van de institutionele omgeving, zoals de wetten die relevant zijn voor de taak die uitgevoerd wordt. Het type werknemer in de publieke sector dat wordt geconfronteerd met anchoring bias of de debiasing interventie lijkt er ook toe te doen. Een volgende stap is het testen van de 'consider-the-opposite' strategie in de praktijk.

Uitgebreide gedragsmodellen kunnen helpen om actieve participatie te begrijpen (deelvraag 3)

We kunnen uitgebreide gedragsmodellen gebruiken, zoals het Theoretical Domains Framework (TDF), om gedrag te begrijpen. In hoofdstuk 4 is dit toegepast om de actieve deelname van politici tijdens online deliberatie te bestuderen. We hebben semi-gestructureerde interviews afgenumen en deze interviews vervolgens geanalyseerd aan de hand van een uitgebreid gedragsmodel: het Theoretical Domains Framework (TDF) (Cane et al., 2012; Michie et al., 2005).

Door de COVID-19-crisis moesten openbare besluitvormende organen online vergaderen. Het is bekend dat online omgevingen leiden tot ongelijkheid vanwege verschillen in niet alleen toegang, maar ook in de manier waarop mensen online actief deelnemen. We weten bovendien dat persoonlijke en contextuele gedragsfactoren deze verschillen kunnen verklaren en tegelijkertijd mogelijkheden bieden voor interventie. Onze resultaten laten zien dat persoonlijke factoren (bijvoorbeeld ideeën over of reageren een doel dient) en contextuele factoren (zoals regels in vergaderingen) actieve deelname beïnvloeden. We raden werknemers binnen het publieke domein aan om ondersteuning aan te bieden om een gebrek aan vaardigheden aan te pakken, in te grijpen op opvattingen over consequenties en passende regels te maken door best practices te delen.

Algoritmen kunnen worden gebruikt voor besluitvorming, afhankelijk van de complexiteit van de taak (deelvraag 4)

Algoritmen kunnen worden gebruikt voor besluitvorming in de publieke sector, maar de toepassingsmogelijkheden zijn afhankelijk van de complexiteit van de taak. We hebben dit onderzocht door twee survey experimenten uit te voeren waarin we hebben getest hoe drie soorten interacties tussen algoritmen en managers een invloed hadden op procedurele rechtvaardigheidspercepties in situaties met een lage en hoge complexiteit. We ontdekten dat wanneer algoritmen onafhankelijk werken, de perceptie van procedurele rechtvaardigheid het meest verandert. De richting van het effect is echter afhankelijk van de complexiteit van de taak. Wanneer een taak relatief eenvoudig is, is procedurele rechtvaardigheid hoger wanneer de beslissing door het algoritme wordt genomen in vergelijking met een manager. Met andere woorden, algoritmen hebben de voorkeur boven managers wanneer taken eenvoudig zijn. Wanneer taken echter complex zijn, vermindert de procedurele rechtvaardigheid. Met andere woorden, managers hebben de voorkeur boven het alleen gebruiken van algoritmen. Dit gebeurt omdat publiek werknemers voor simpele praktijken waarde hechten aan gelijke behandeling van werknemers op basis van regels. Wanneer praktijken echter complex zijn, waarderen ze de beslissingsruimte van het management en geloven ze niet dat het algoritme alle variabelen kan meenemen.

We ontdekten ook dat het toevoegen van een algoritme aan publieke besluitvorming de procedurele rechtvaardigheid kan vergroten, maar de effecten zijn wederom afhankelijk van de situatie. Wanneer taken eenvoudig zijn, heeft hybride besluitvorming de voorkeur boven een manager die alleen beslist. Wanneer praktijken complex zijn, wordt een

combinatie als rechtvaardiger gezien, maar alleen wanneer een combinatie wordt vergeleken met besluitvorming door een publieke manager. Anders wordt besluitvorming door een manager en hybride besluitvorming gezien als even procedureel rechtvaardig. Dit betekent dat de aanvaardbaarheid van veranderende besluitvormingsprocessen moet worden bekeken in relatie tot wie of wat momenteel de beslissingen neemt.

DRIE HOOFDCONCLUSIES

Op basis van de antwoorden op onze deelvragen trekken we drie hoofdconclusies.

Gedragswetenschappen kunnen besluitvorming ondersteunen door middel van nudges, debiasing, uitgebreide analyses en specifieke psychologische theorieën

Onze eerste hoofdconclusie is dat we gedragswetenschappen op vier manieren kunnen gebruiken om besluitvorming in de publieke sector te ondersteunen: door nudging, door debiasing interventies toe te passen, door uitgebreide gedragsanalyses te gebruiken die interventies informeren, en door gebruik te maken van specifieke psychologische theorieën. Ten eerste kunnen we nudges gebruiken om de besluitvorming van werknemers in de publieke sector te ondersteunen. Er zijn veel verschillende toepassingen van nudges mogelijk. In hoofdstuk 2 zagen we dat zorgprofessionals genudged worden naar meer evidence-based medicine. Dit gebeurde onder meer door middel van geheugensteuntjes, door gebruik te maken van feedback over toppresteerders en door voorgestelde alternatieven te tonen. Dit kan worden toegepast op andere werknemers in de publieke sector dan professionals in de gezondheidszorg. Recent bewijs heeft bijvoorbeeld aangetoond dat kleine financiële prikkels en geheugensteuntjes invloed hadden op het verstrekken van informatie aan het ministerie van Onderwijs in Denemarken (Andersen & Hvidman, 2021). We willen ook het potentieel voor digitale toepassingen benadrukken, via digitale systemen. Digitale omgevingen hebben aanpasbare keuzearchitecturen die in de toekomst waarschijnlijk belangrijker zullen worden. Dit heeft te maken met de digitalisering van veel overheidsdiensten (Bovens & Zoidis, 2002). Daarbij kunnen we gebruik maken van onderzoek naar digitale nudging of e-nudging, wat 'het gebruik van gebruikersinterface-ontwerpelementen is om het gedrag van mensen in digitale keuzeomgevingen te sturen' (Weinmann et al., 2016, p.1).

Ten tweede kunnen biases een invloed hebben op de besluitvorming door werknemers in de publieke sector (Battaglio et al., 2018). In hoofdstuk 3 hebben we gezien dat anchoring bias, wat de neiging is om hoeveelheden in te schatten met behulp van een initiële waarde (Tversky & Kahneman, 1974), de besluitvorming van werknemers in de publieke sector beïnvloedt. Dit effect is robuust in verschillende institutionele contexten. In hoofdstuk 3 hebben we ook gezien dat we kunnen ingrijpen op biases door een goedkope en makkelijke debiasing interventie toe te passen. Bij het ontwerpen van debiasing interventies moeten we er rekening mee houden dat de mechanismen van de bias overeenkomen met de debiasing interventie (Croskerry et al., 2013a; Larrick, 2004). Zo zagen we in hoofdstuk 3 dat anchoring bias een tunnelvisie creëert waarin alternatieven ontoegankelijk worden in de geest (Furnham & Boo, 2011). Als we de consider-the-opposite strategie gebruiken, doorbreken we deze tunnelvisie (Mussweiler et al., 2000). Debiasing-strategieën kunnen

ook worden toegepast op andere biases waarvan we weten dat een effect hebben op de besluitvorming in de publieke sector, zoals confirmation bias. Voor andere biases hebben we mogelijk andere technieken nodig (Hallsworth et al., 2018).

Ten derde kunnen uitgebreide analyses helpen om problemen vanuit een gedragsperspectief te begrijpen. In hoofdstuk 4 hebben we het Theoretical Domains Framework (TDF) gebruikt om individuele verschillen in actieve participatie tijdens online beraadslagingen van te analyseren (Michie et al., 2014; Michie et al., 2005). We ontdekten dat zowel persoonlijke als contextuele factoren belangrijk waren bij het verklaren van actieve deelname aan online beraadslagingen. Deze analyse kan vervolgens worden gebruikt om interventies te maken die de besluitvorming kunnen ondersteunen. In hoofdstuk 4 ontdekten we bijvoorbeeld dat doelen belangrijk waren. Doelen kunnen worden verandert door interventies zoals het maken van een actieplan (Michie et al., 2014). Andere factoren kunnen echter andere interventies vereisen. Dit soort analyses wordt zelden gebruikt in bestuurskundig onderzoek, maar het gebruik ervan kan inzicht geven in de oorzaken van gedrag in de publieke sector (Meder et al., 2018; Olejniczak et al., 2020; Weaver, 2015).

Ten vierde kunnen we het gebruik van specifieke psychologische theorieën in de bestuurskunde voortzetten en uitbreiden. In hoofdstuk 5 gebruiken we bijvoorbeeld procedurele rechtvaardigheid om te onderzoeken hoe de betrokkenheid van algoritmen bij managementbeslissingen percepties verandert voor taken die verschillen in complexiteit. Wetenschappers hebben dit potentieel erkend en hebben verschillende onderzoeken gedaan naar verschillende aspecten van gedragswetenschappen en de bestuurskunde. Bekende voorbeelden zijn onderzoek naar representatieve bureaucratie, public service motivation en beleidsvervreemding. Momenteel breiden wetenschappers echter de verbindingen tussen gedragswetenschappen en openbaar bestuur uit. Recent werk van Aleksovska (2021) heeft bijvoorbeeld aangetoond dat verantwoording kan worden begrepen vanuit een gedragsperspectief.

Gedragswetenschappen kunnen worden gecontextualiseerd door rekening te houden met institutionele en situationele instellingen

Wanneer we werken met gedragswetenschappen in de publieke sector, moeten we rekening houden met de institutionele en situationele omgeving waarin gedrag plaatsvindt. Ten eerste kan de institutionele omgeving een verschil maken. In hoofdstuk 3 hebben we bijvoorbeeld een experiment in het Verenigd Koninkrijk (VK) gerepliceerd dat eerder in Italië is uitgevoerd. We vonden dat alhoewel de richting van het effect hetzelfde bleef, de effectgroottes en de minima- en maxima-gemiddelden substantieel verschilden. Dit verschil was groter voor het experiment met betrekking tot het stellen van doelen voor responsiviteit naar burgers dan voor ons experiment over prestatiebeoordeling. In hoofdstuk 3 hebben we uitgelegd dat dit veroorzaakt kan worden door de institutionele omgeving die relevant is voor deze twee verschillende praktijken. De verschillen tussen het VK en Italië kunnen bijvoorbeeld worden verklaard door verschillende wetten met betrekking tot de responstijd voor burgers die verschillen in het VK en Italië. Als zodanig kan het effect van biases worden beïnvloed door instituties.

In hoofdstuk 4 zagen we ook dat instituties keuzearchitectuur kunnen creëren. Dit gebeurde door de invoering van gedragsregels. Meer specifiek werden, als reactie op de nadelen van digitaal overleg, regels gemaakt die ongunstige keuze-architecturen creëerden. Gemeenten hebben regels opgesteld over wanneer en hoe er op elkaar gereageerd moet worden. Zo werd aan politici gevraagd te reageren nadat iedereen zijn of haar bijdrage had geleverd. Dit zorgde er echter voor dat mensen barrières ondervonden om te reageren: ze konden zich niet alles herinneren wat er was gezegd en hadden vaak het gevoel dat het reageren 'mosterd na de maaltijd' was. Als zodanig creëerden de vastgestelde regels psychologische barrières die moeilijk te overwinnen waren. Dit betekent dat we bij het bestuderen van keuzearchitectuur verder moeten gaan dan alleen de tastbare omgeving, maar ook rekening moeten houden met de instituties. Dit kan met name relevant zijn binnen organisaties die onvermijdelijk uit instituties bestaan (March & Olsen, 2019).

Ten tweede toonden we aan dat de situationele omgeving ertoe doet. Dit had vooral te maken met de complexiteit van de taak. In hoofdstuk 2 werd duidelijk dat nudges die gericht waren op relatief onbetwiste onderwerpen zoals handhygiëne het meest effectief waren. Nudges die gericht waren op complexere onderwerpen waren minder effectief. Dit kan gebeuren omdat nudges vaak niet gecontextualiseerd zijn, waardoor ze ongeschikt zijn voor de complexe realiteit van de medische praktijk. Dit wordt bevestigd door Huis in 't Veld et al. (2020) die stellen dat medische professionals vinden dat de doelen van nudges zo objectief mogelijk moeten zijn. Het gebruik van algoritmen om nudges te contextualiseren zou een optie kunnen zijn. In hoofdstuk 5 hebben we onderzocht wat er gebeurt als algoritmen worden gebruikt voor besluitvorming in het publieke domein. We ontdekten dat voor simpele praktijken, algoritmen als geschikt worden beschouwd voor besluitvorming, terwijl voor complexe praktijken een manager of een hybride vorm van besluitvorming de voorkeur heeft. Een algoritme zou in die gevallen een suggestie kunnen geven aan een manager, wat kan worden gezien als nudge (Vaccaro & Waldo, 2019). Dit houdt in dat het gebruik van nudges voor zaken met een hoge complexiteit een optie is.

In hoofdstuk 5 zagen we ook dat complexiteit cognitieve processen directer beïnvloedde. Op het moment dat online deliberaties bijvoorbeeld complexer worden omdat er meer actoren aanwezig waren, was het voor mensen moeilijker om het overzicht te bewaren. Dit zorgde ervoor dat politici zich minder bewust waren van wie wat zei in beraadslagingen en op wie ze wilden reageren. Dit betekent dat complexiteit bestaande cognitieve problemen versterkt. Dit betekent dat complexiteit een invloed heeft die verder gaat dan heuristieken en bias.

Gedragsinterventies kunnen worden gepersonaliseerd door rekening te houden met individuele verschillen

Onze derde hoofdconclusie is dat we gedragsinterventies kunnen personaliseren door rekening te houden met individuele verschillen. We kunnen dit op twee manieren doen. Ten eerste kunnen we onderscheid maken tussen verschillende subgroepen bij het

toepassen van nudges en debiasing interventies. In hoofdstuk 3 zagen we bijvoorbeeld dat publieke managers en medewerkers verschillend reageerden op onze debiasing interventie. Dit houdt in dat managers en medewerkers op een andere manier beslissingen te nemen, en niet alle groepen hetzelfde beïnvloed worden door debiasing interventies. Naast het onderscheiden van verschillende soorten beroepen, kunnen we onderscheid maken tussen verschillende soorten professionals. In hoofdstuk 2 ontdekten we dat het type professional ook een verschil kan maken in de manier waarop met nudges wordt omgegaan. Jongere professionals kunnen bijvoorbeeld gevoeliger zijn voor nudges dan oudere professionals. Dit geeft aan dat professionaliteit en gedragsinterventies, zoals nudges, op gespannen voet met elkaar kunnen staan.

Ten tweede kunnen we gedragsinterventies afstemmen op groepen en individuele behoeften. Dit gaat verder dan het aanpassen van nudges en debiasing interventies aan verschillende beroepsmatige en professionele subgroepen. In plaats daarvan kunnen uitgebreide gedragsanalyses psychologische factoren identificeren voorbij heuristieken en biases. In het verlengde hiervan kan dit leiden tot de keuze voor andere soorten gedragsinterventies. In hoofdstuk 4 hebben we laten zien dat individuele verschillen veroorzaakt worden door persoonlijke psychologische factoren zoals vaardigheden, overtuigingen over of reageren een doel dient en de doelen die een persoon heeft. Deze individuele verschillen kunnen per subgroep verschillen. In hoofdstuk 4 zagen we bijvoorbeeld dat ouderen, bijvoorbeeld bij digitale beraadslagingen, vaak minder digitale vaardigheden hebben en mogelijk meer ondersteuning nodig hebben. Algoritmen kunnen oplossingen bieden voor beide personalisatiemogelijkheden. Dit kan door differentiatie op subgroepen, maar ook door het personaliseren van interventies (Ebert & Freibichler, 2017; Michie et al., 2017).

S

IMPLICATIES

Deze drie hoofdconclusies hebben implicaties voor zowel de wetenschap als de praktijk. Voor de wetenschap raden we aan om onderscheid te maken tussen een beperkte en een brede definitie van gedragsbestuurkunde (Ewert, 2020; Gofen et al., 2021). De beperkte definitie verwijst naar het gebruik van inzichten over heuristieken en biases en met name het afwijken van rationele modellen van besluitvorming. De brede definitie houdt rekening met cognitieve processen in het algemeen die de besluitvorming kunnen beïnvloeden. Ten tweede moeten we nauwkeuriger nadenken over hoe we gedragswetenschappen specifiek kunnen gebruiken in publieke contexten (Ewert et al., 2021; Feitsma & Whitehead, 2019; Jilke et al., 2019). Dit proefschrift laat zien hoe de institutionele en situationele omgeving belangrijk is voor gedragsbestuurskundig onderzoek. Ten derde moeten we onderscheid maken tussen verschillende werknemers in de publieke sector. Werknemers in de publieke sector zijn divers in termen van beroepen, professionaliteit en andere persoonlijke factoren. Daarom moeten we verder gaan dan alleen one-size-fits-all benaderingen als we werknemers in de publieke sector willen beïnvloeden. In plaats daarvan kunnen we interventies differentiëren en personaliseren. Ten vierde kunnen we verschillende methoden gebruiken om gedragsfactoren te onderzoeken die van

invloed zijn op de besluitvorming in het openbaar bestuur naast het survey-experiment. In plaats daarvan kunnen we gebruik maken van systematische (scoping) reviews, replicaties en extensies, 'mixed methods' en kwalitatieve methoden. Ten vijfde kunnen gedragswetenschappen expliciter worden gekoppeld aan nieuwe onderwerpen die te maken hebben met digitalisering. Deze onderwerpen zijn op dit moment zeer relevant, aangezien we toewerken naar bureaucratie op 'systeem- en screen level' (Zouridis et al., 2020). Algoritmen kunnen bovendien worden begrepen door gedragswetenschappen en bieden heterogeniteit in interventies op basis van contextuele en individuele verschillen.

Voor de praktijk is de eerste aanbeveling om te zoeken naar mogelijkheden om binnen publieke organisaties om nudges toe te passen om de besluitvorming van werknemers in de publieke sector te ondersteunen. Dit kan door het aanpassen van fysieke omgevingen zoals kantoorgebouwen, op fysieke objecten zoals formulieren, van menselijke interacties maar ook in digitale omgevingen. Onze tweede aanbeveling is om je bewust te zijn van het effect van biases en waar nodig debiaserende interventies toe te passen. Deze biases kunnen in bijna alle stadia van het beleidsproces aanwezig zijn. Ze kunnen aanwezig zijn in hoe mensen informatie interpreteren, hoe mensen met elkaar omgaan binnen groepen, maar ook in hoe beleid wordt geïmplementeerd (zie Battaglio et al., 2018; Hallsworth et al., 2018; Moseley & Thomann, 2021 voor meer voorbeelden van biases in publieke besluitvorming). Voor het ontwerpen van een debiasing interventie is het essentieel om het mechanisme te begrijpen waardoor de bias werkt. In ons werk was de 'consider-the-opposite' strategie bijvoorbeeld effectief om anchoring bias tegen te gaan, maar andere biases hebben mogelijk andere debiasing-interventies nodig. Onze derde aanbeveling is om andere psychologische modellen toe te passen om nieuwe besluitvormingssituaties of problemen in bestaande besluitvormingssituaties te analyseren. Het gebruik van een veelzijdig algemeen model dat kan worden gekoppeld aan specifieke gedragsinterventies zoals het Theoretical Domains Framework of het COM-B-model kan een goed uitgangspunt zijn (Michie et al., 2014). Onze vierde aanbeveling is om rekening te houden met de institutionele en situationele omgeving, evenals met persoonlijke kenmerken van werknemers in de publieke sector.

CONCLUSIE

We concluderen dat wetenschappers en mensen uit de praktijk de introductie van heuristieken het beste kunnen hanteren door conceptueel een beperkte en brede definitie van gedragswetenschappen te erkennen. Empirisch kunnen we nudges, debiasing interventies, uitgebreide gedragsmodellen en specifieke psychologische theorieën gebruiken. Daarbij moeten we de institutionele en situationele omgeving in beschouwing nemen. We moeten ook onderscheid maken tussen werknemers in de publieke sector en interventies personaliseren. Algoritmen kunnen helpen bij het contextualiseren en personaliseren van gedragsbestuurskunde. Verder onderzoek is nodig om het potentieel van het gebruik van algoritmen in combinatie met gedragswetenschappen in de bestuurskunde te verkennen. Als zodanig kunnen we heuristieken hanteren om de besluitvorming in de context van de publieke sector te begrijpen en te ondersteunen.

S

REFERENCES

REFERENCES

- 't Hart, P. (1998). Preventing groupthink revisited: Evaluating and reforming groups in government. *Organizational Behavior and Human Decision Processes*, 73(2-3), 306–326.
- Adame, B. J. (2016). Training in the mitigation of anchoring bias: A test of the consider-the-opposite strategy. *Learning and Motivation*, 53, 36–48.
- Adhikari, D. (2016). Exploring the differences between social and behavioral science. *Behavioral Development Bulletin*, 21(2), 128–135.
- Albrecht, S. (2006). Whose voice is heard in online deliberation?: A study of participation and representation in political debates on the internet. *Information Communication and Society*, 9(1), 62–82.
- Aleksovska, M., Schillemans, T., & Grimmelikhuijsen, S. (2019). Lessons from five decades of experimental and behavioral research on accountability: A systematic literature review. *Journal of Behavioral Public Administration*, 2(2), 1–18.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9–10), 1082–1095.
- Allen, D., & Harkins, K. J. (2005). Too much guidance? *Lancet*, 365(9473), 1768.
- 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.
- Anastasopoulos, L. J., & Whitford, A. B. (2019). Machine Learning for Public Administration Research, With Application to Organizational Reputation. *Journal of Public Administration Research and Theory*, 29(3), 491–510.
- Andersen, S. C., & Hvidman, U. (2021). Can Reminders and Incentives Improve Implementation Within Government? Evidence from a Field Experiment. *Journal of Public Administration Research and Theory*, 31(1), 234–249.
- Anderson, C. A. (1982). Inoculation and Counterexplanation: Debiasing Techniques in the Perseverance of Social Theories. *Social Cognition*, 1(2), 126–139.
- Ansher, C., Ariely, D., Nagler, A., Rudd, M., Schwartz, J., & Shah, A. (2014). Better Medicine by Default. *Medical Decision Making*, 34(2), 147–158.
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalivé, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21(6), 1086–1120.
- Appelbaum, E., Bailey, T., Berg, P., & Kalleberg, A. (2000). Manufacturing advantage: Why high-performance work systems pay off. Cornell University Press.
- Appelt, K. C., Milch, K. F., Handgraaf, M. J., & Weber, E. U. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, 6(3), 252–262.
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes*, 37(1), 93–110.
- Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, 73(2), 305–307.
- Aronson, E. (2018). Social psychology. Pearson Educated Limited.
- Atkins, L., Francis, J., Islam, R., O'Connor, D., Patey, A., Ivers, N., Foy, R., Duncan, E. M., Colquhoun, H., Grimshaw, J. M., Lawton, R., & Michie, S. (2017). A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implementation Science*, 12(1), 1–18.
- Atkins, L., Sallis, A., Chadborn, T., Shaw, K., Schneider, A., Hopkins, S., Bunten, A., Michie, S., & Lorencatto, F. (2020). Reducing catheter-associated urinary tract infections: A systematic review of barriers and facilitators and strategic behavioural analysis of interventions. *Implementation Science*, 15(1), 1–22.

- Bächtiger, A., Dryzek, J. S., Mansbridge, J., & Warren, M. (2018). Deliberative Democracy. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. Warren (Eds.), *The Oxford Handbook of Deliberative Democracy* (pp. xxii–32). Oxford University Press.
- Bächtiger, A., Niemeyer, S., Neblo, M., & Steenbergen, M. (2010). Disentangling diversity in deliberative democracy: Competing theories, their blind spots and complementarities. *The Journal of Political Philosophy*, 18(1), 32–63.
- Baekgaard, M., Christensen, J., Dahlmann, C. M., Mathiasen, A., & Petersen, N. B. G. (2019). The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians. *British Journal of Political Science*, 49(3), 1117–1140.
- Barabas, J. (2004). How Deliberation Affects Policy Opinions. *American Political Science Review*, 98(4), 687–701.
- Battaglio, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2018). Behavioral Public Administration ad fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review*, 79(3), 1–17.
- Bauer, M., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology*, 13, 3–32.
- Baviskar, S., & Winter, S. C. (2017). Street-Level Bureaucrats as Individual Policymakers: The Relationship between Attitudes and Coping Behavior toward Vulnerable Children and Youth. *International Public Management Journal*, 20(2), 316–353.
- Beauvais, E., & Beauvais, E. (2018). Deliberation and Equality. In *The Oxford Handbook of Deliberative Democracy* (pp. 143–155). Oxford University Press.
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13.
- Beidas, R. S., Becker-Haimes, E. M., Adams, D. R., Skriner, L., Stewart, R. E., Wolk, C. B., Buttenheim, A. M., Williams, N. J., Inacker, P., Richey, E., & Marcus, S. C. (2017). Feasibility and acceptability of two incentive-based implementation strategies for mental health therapists implementing cognitive-behavioral therapy: a pilot study to inform a randomized controlled trial. *Implementation Science*, 12(1), 148.
- Belardinelli, P., Bellé, N., Sicilia, M., & Steccolini, I. (2018). Framing Effects under Different Uses of Performance Information: An Experimental Study on Public Managers. *Public Administration Review*, 78(6), 841–851.
- Bellé, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive Biases in Performance Appraisal: Experimental Evidence on Anchoring and Halo Effects With Public Sector Managers and Employees. *Review of Public Personnel Administration*, 37(3), 275–294.
- Bellé, N., Cantarelli, P., & Belardinelli, P. (2018). Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions. *Public Administration Review*, 78(6), 828–840.
- Benson, B. (2016). Cognitive bias cheat sheet – Better Humans. Retrieved January, 8, 2019, from <https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18>
- Bhanot, S. P., & Linos, E. (2020). Behavioral Public Administration: Past, Present, and Future. *Public Administration Review*, 80(1), 168–171.
- Bindels, R., Hasman, A., Kester, A., Talmon, J. L., de Clercq, P. A., & Winkens, R. A. G. (2003). The efficacy of an automated feedback system for general practitioners. *Informatics in Primary Care*, 11(2), 69–74.
- Binnenlands Bestuur. (2020). Eerste Kamer akkoord met digitale besluitvorming. Retrieved April, 30, 2020, from <https://www.binnenlandsbestuur.nl/bestuur-en-organisatie/nieuws/groen-licht-voor-digitale-besluitvorming.12862290.lynkx>
- Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). "It's reducing a human being to a percentage"; perceptions of justice in algorithmic decisions. *Conference on Human Factors in Computing Systems - Proceedings*, 1–14.

References

- Birken, S. A., Powell, B. J., Presseau, J., Kirk, M. A., Lorencatto, F., Gould, N. J., Shea, C. M., Weiner, B. J., Francis, J. J., Yu, Y., Haines, E., & Damschroder, L. J. (2017). Combined use of the Consolidated Framework for Implementation Research (CFIR) and the Theoretical Domains Framework (TDF): A systematic review. *Implementation Science*, 12(1), 1–14.
- Birnbach, D., King, D., Vlaev, I., Rosen, L. F., & Harvey, P. D. (2013). Impact of environmental olfactory cues on hand hygiene behaviour in a simulated hospital environment: a randomized study. *Journal of Hospital Infection*, 85(1), 79–81.
- Bishu, S. G., & Kennedy, A. R. (2020). Trends and Gaps: A Meta-Review of Representative Bureaucracy. *Review of Public Personnel Administration*, 40(4), 559–588.
- Boers, E. (2017). BOB bedreigt de raadsvergadering - Gemeente.nu. Retrieved December, 21, 2020, from <https://www.gemeente.nu/raadsledennieuws/bob-bedreigt-de-raadsvergadering/>
- Bogliacino, F., Codagnone, C., & Veltri, G. A. (2016). An introduction to the special issue on “the behavioural turn in public policy: new evidence from experiments.” *Economia Politica*, 33(3), 323–332.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons.
- Bourdeaux, C. P., Davies, K. J., Thomas, M. J. C., Bewley, J. S., & Gould, T. H. (2014). Using ‘nudge’ principles for order set design: a before and after evaluation of an electronic prescribing template in critical care. *BMJ Quality & Safety*, 23(5), 382–388.
- Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014. *International Journal of Public Sector Management*, 29(2), 110–131.
- Bovens, M., & Zouridis, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review*, 62(2), 174–184.
- Bozeman, B. (1993). A Theory Of Government “Red Tape.” *Journal of Public Administration Research and Theory*, 3(3), 273–304.
- Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, 28(2), 151–155.
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van ’t Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224.
- Breeman, G., Noort, W. van, & Rutgers, M. (2016). *De bestuurlijke kaart van Nederland: het openbaar bestuur en zijn omgeving in nationaal en internationaal perspectief* (6th ed.). Coutinho.
- Brown, M. (2018). Deliberation and Representation. In *The Oxford Handbook of Deliberative Democracy* (pp. 170–186). Oxford University Press.
- Buising, K. L., Thursky, K. A., Black, J. F., MacGregor, L., Street, A. C., Kennedy, M. P., & Brown, G. V. (2008). Improving antibiotic prescribing for adults with community acquired pneumonia: Does a computerised decision support system achieve more than academic detailing alone? – a time series analysis. *BMC Medical Informatics and Decision Making*, 8(1), 1–10.
- Burton, J. W., Stein, M., & Jensen, T. B. (2019). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239.
- 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.
- Busch, P. A., Henriksen, H. Z., & Sæbø, Ø. (2018). Opportunities and challenges of digitized discretionary practices: a public service worker perspective. *Government Information Quarterly*, 35(4), 547–556.
- Butler, D. M., & Broockman, D. E. (2011). Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science*, 55(3), 463–477.

- Cane, J., O'Connor, D., & Michie, S. (2012). Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation Science*, 7(1), 1–17.
- Cantarelli, P., Bellé, N., & Belardinelli, P. (2018). Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions. *Review of Public Personnel Administration*, 40(1), 56–81.
- Carrigan, C., Pandey, S. K., & Ryzin, G. G. Van. (2020). Pursuing Consilience: Using Behavioral Public Administration to Connect Research on Bureaucratic Red Tape, Administrative Burden, and Regulation. *Public Administration Review*, 80(1), 46–52.
- Casper, E. S. (2008). Using implementation intentions to teach practitioners: changing practice behaviors via continuing education. *Psychiatric Services*, 59(7), 747–752.
- Centraal Bureau voor de Statistiek. (2021). Gemeentelijke indeling op 1 januari 2021. Retrieved April 30, 2020, from <https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen-per-jaar/indeling-per-jaar/gemeentelijke-indeling-op-1-januari-2020>
- Chaiken, S., & Trope, Y. (1999). Dual-process theories in social psychology. Guilford Press.
- Chan, B. P., Homa, K., & Kirkland, K. B. (2013). Effect of Varying the Number and Location of Alcohol-Based Hand Rub Dispensers on Usage in a General Inpatient Medical Unit. *Infection Control & Hospital Epidemiology*, 34(09), 987–989.
- Chapman, G. B., & Johnson, E. J. (1999). Anchoring, Activation, and the Construction of Values. *Organizational Behavior and Human Decision Processes*, 79(2), 115–153.
- Chapman, G., Milkman, K. L., Rand, D., Rogers, T., & Thaler, R. H. (2021). Nudges and choice architecture in organizations: New frontiers. *Organizational Behavior and Human Decision Processes*, 163(March 2021), 1–3.
- Chen, Z. J., Vogel, D., & Wang, Z. H. (2016). How to satisfy citizens? Using mobile government to reengineer fair government processes. *Decision Support Systems*, 82(February 2016), 47–57.
- Chen, Z., & Kemp, S. (2015). Anchoring Effects in Simulated Academic Promotion Decisions: How the Promotion Criterion Affects Ratings and the Decision to Support an Application. *Journal of Behavioral Decision Making*, 28(2), 137–148.
- Cheng, F.-F., Wu, C.-S., & Lin, H.-H. (2014). Reducing the influence of framing on internet consumers' decisions: The role of elaboration. *Computers in Human Behavior*, 37(August 2014), 56–63.
- Cheung, A., Weir, M., Mayhew, A., Kozloff, N., Brown, K., & Grimshaw, J. (2012). Overview of systematic reviews of the effectiveness of reminders in improving healthcare professional behavior. *Systematic Reviews*, 1(36), 1–8.
- Christensen, J. (2018). Let's look at the facts: An investigation of psychological biases in policymakers' interpretation of policy-relevant information. Aarhus University.
- Christensen, J., Aarøe, L., Baekgaard, M., Herd, P., & Moynihan, D. P. (2020). Human Capital and Administrative Burden: The Role of Cognitive Resources in Citizen-State Interactions. *Public Administration Review*, 80(1), 127–136.
- Christensen, J., & Moynihan, D. P. (2020). Motivated reasoning and policy information: politicians are more resistant to debiasing interventions than the general public. *Behavioural Public Policy*, 1–22.
- Citron, D., & Pasquale, F. (2014). The scored society: due process for automated predictions. *Washington Law Review*, 89(1), 1–34.
- Colquitt, J. A. (2001). On the Dimensionality of Organizational Justice: A Construct Validation of a Measure. *Journal of Applied Psychology*, 86(3), 386–400.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research. *Journal of Applied Psychology*, 86(3), 425–445.

References

- Cooper, H. (2010). Research Synthesis and Meta-Analysis: A Step-by-Step Approach (4th ed.). SAGE.
- Crawford, K., & Schultz, J. (2014). Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms Recommended Citation. *Boston College Law Review*, 55(2), 1–29.
- Creedon, S. A. (2005). Healthcare workers' hand decontamination practices: Compliance with recommended guidelines. *Journal of Advanced Nursing*, 51(3), 208–216.
- Croskerry, P., Singhal, G., & Mamede, S. (2013a). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Quality & Safety*, 22, ii58–ii64.
- Croskerry, P., Singhal, G., & Mamede, S. (2013b). Cognitive debiasing 2: impediments to and strategies for change. *BMJ Quality & Safety*, 22, ii65–ii72.
- Cummings, K. M., Frisof, K. B., Long, M. J., & Hrynkiewich, G. (1982). The effects of price information on physicians' test-ordering behavior. *Ordering of diagnostic tests*. *Medical Care*, 20(3), 293–301.
- Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology*, 29(3), 245–268.
- Davis, R., Campbell, R., Hildon, Z., Hobbs, L., & Michie, S. (2014). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review*, 9(3), 323–344.
- de Boer, N. (2020). How Do Citizens Assess Street-Level Bureaucrats' Warmth and Competence? A Typology and Test. *Public Administration Review*, 80(4), 532–542.
- Dekker, R., & Bekkers, V. (2015). The contingency of governments' responsiveness to the virtual public sphere: A systematic literature review and meta-synthesis. *Government Information Quarterly*, 32(4), 496–505.
- Denton, G. D., Smith, J., Faust, J., & Holmboe, E. (2001). Comparing the efficacy of staff versus housestaff instruction in an intervention to improve hypertension management. *Academic Medicine*, 76(12), 1257–1260.
- Desouza, K. C., & Jacob, B. (2017). Big Data in the Public Sector: Lessons for Practitioners and Scholars. *Administration & Society*, 49(7), 1043–1064.
- Dexter, P., Perkins, S., Maharry, K., & Jones, K. (2004). Inpatient computer-based standing orders vs physician reminders to increase influenza and pneumococcal vaccination rates: a randomized trial. *JAMA*, 292(19), 2366–2371.
- Diakopoulos, N. (2014). Algorithmic Accountability Reporting: On the Investigation of Black Boxes. *Tow Center for Digital Journalism*.
- Diritto di Sapere. (2017). Ignoranza di Stato - Rapporto sull'applicazione del Foia italiano -. Retrieved November 14, 2019, from <https://blog.dirittodisapere.it/rapporto-foia/>
- Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindsphere way. *Journal of Economic Psychology*, 33(1), 264–277.
- Donoghue, J. (2017). The Rise of Digital Justice: Courtroom Technology, Public Participation and Access to Justice. *The Modern Law Review*, 80(6), 995–1025.
- Dooren, W. Van, Bouckaert, G., & Halligan, J. (2015). Performance Management in the Public Sector. Routledge.
- Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral Vision Expertise in Real World Contexts. *Organization Studies*, 26(5), 779–792.
- Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (2011). Cambridge Handbook of Experimental Political Science (J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (eds.)). Cambridge University Press.
- Dudley, S. E., & Xie, Z. (2020). Designing a Choice Architecture for Regulators. *Public Administration Review*, 80(1), 151–156.

- Dustan, A., Maldonado, S., & Hernandez-Agramonte, J. M. (2019). Motivating Bureaucrats With Non-Monetary Incentives When State Capacity Is Weak: Evidence From Large-Scale Field Experiments in Peru. *SSRN Electronic Journal*.
- Dwivedi, Y. K., Rana, N. P., Janssen, M., Lal, B., Williams, M. D., & Clement, M. (2017). An empirical validation of a unified model of electronic government adoption (UMEGA). *Government Information Quarterly*, 34(2), 211–230.
- Ebert, P., & Freibichler, W. (2017). Nudge management: applying behavioural science to increase knowledge worker productivity. *Journal of Organization Design*, 6(1), 4.
- Engin, Z., & Treleaven, P. (2019). Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies. *The Computer Journal*, 62(3), 448–460.
- Englich, B., Mussweiler, T., & Strack, F. (2006). Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin*, 32(2), 188–200.
- Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors. In *Psychological Science* (Vol. 12, Issue 5).
- Erasmus, V., Kuperus, M. N., Richardus, J. H., Vos, M. C., Oenema, A., & van Beeck, E. F. (2010). Improving hand hygiene behaviour of nurses using action planning: A pilot study in the intensive care unit and surgical ward. *Journal of Hospital Infection*, 76(2), 161–164.
- Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454–459.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition. *Perspectives on Psychological Science*, 8(3), 223–241.
- Evans, T., & Hupe, P. (2019). Discretion and the quest for controlled freedom. Palgrave Macmillan.
- Ewert, B. (2020). Moving beyond the obsession with nudging individual behaviour: Towards a broader understanding of Behavioural Public Policy. *Public Policy and Administration*, 35(3), 337–360.
- Ewert, B., Loer, K., & Thomann, E. (2021). Beyond nudge: advancing the state-of-the-art of behavioural public policy and administration. *Policy and Politics*, 49(1), 3–23.
- Fairclough, I. (2019). Deontic power and institutional contexts: The impact of institutional design on deliberation and decision-making in the UK fracking debate. *Journal of Argumentation in Context*, 8(1), 136–171.
- Faulkner, N., Borg, K., Bragge, P., Curtis, J., Ghafoori, E., Goodwin, D., Jorgensen, B. S., Jungbluth, L., Kneebone, S., Smith, L., Wright, B., & Wright, P. (2018). The INSPIRE Framework: How Public Administrators Can Increase Compliance with Written Requests Using Behavioral Techniques. *Public Administration Review*, 79(1), 125–135.
- Favero, N., Meier, K. J., & O'Toole, L. J. (2016). Goals, Trust, Participation, and Feedback: Linking Internal Management With Performance Outcomes. *Journal of Public Administration Research and Theory*, 26(2), 327–343.
- Feitsma, J. (2018). Brokering behaviour change: the work of behavioural insights experts in government. *Policy & Politics*, 47(1), 37–56.
- Feitsma, J. (2019). Inside the Behavioural State. Eleven International Publishing.
- Feitsma, J., & Whitehead, M. (2019). Bounded interdisciplinarity: critical interdisciplinary perspectives on context and evidence in behavioural public policies. *Behavioural Public Policy*, 1–27.
- Feldstein, A., Smith, D., Perrin, N., & Yang, X. (2006). Improved therapeutic monitoring with several interventions: a randomized trial. *Archives of Internal Medicine*, 166(17), 1848–1854.

References

- Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories. *Perspectives on Psychological Science*, 7(6), 555–561.
- Field, M. J., & Lohr, K. N. (1990). Clinical Practice Guidelines: Directions for a New Program. National Academy Press.
- Field, T. S., Rochon, P., Lee, M., Gavendo, L., Baril, J. L., & Gurwitz, J. H. (2009). Computerized clinical decision support during medication ordering for long-term care residents with renal insufficiency. *Journal of the American Medical Informatics Association*, 16(4), 480–485.
- Filippi, A., Sabatini, A., Badioli, L., Samani, F., Mazzaglia, G., Catapano, A., & Cricelli, C. (2003). Effects of an automated electronic reminder in changing the antiplatelet drug-prescribing behavior among Italian general practitioners in diabetic patients: an. *Diabetes Care*, 26(5), 1496–1500.
- Finck, M. (2019). Smart contracts as a form of solely automated processing under the GDPR. *International Data Privacy Law*, 9(2), 78–94.
- Flanagan, J. R., Doebbeling, B. N., Dawson, J., & Beekmann, S. (1999). Randomized Study of Online Vaccine Reminders in Adult Primary Care. *Proceedings of the AMIA Symposium*, 755–759.
- Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, 12(2), 219–245.
- Fogarty, A. W., Sturrock, N., Premji, K., & Prinsloo, P. (2013). Hospital clinicians' responsiveness to assay cost feedback: a prospective blinded controlled intervention study. *JAMA Internal Medicine*, 173(17), 1654–1655.
- Förberg, U., Unbeck, M., Wallin, L., Johansson, E. M., Petzold, M., Ygge, B.-M., Ehrenberg, A., Petzold, A., Ygge, B. B. M., & Ehrenberg, A. (2016). Effects of computer reminders on complications of peripheral venous catheters and nurses' adherence to a guideline in paediatric care—a cluster randomised study. *Implementation Science*, 11(1), 10.
- Friess, D., & Eilders, C. (2015). A Systematic Review of Online Deliberation Research. *Policy & Internet*, 7(3), 319–339.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42.
- Gawronski, B., & Creighton, L. A. (2013). Dual Process Theories. In *The Oxford Handbook of Social Cognition* (pp. 282–312). Oxford University Press.
- Gemeente Amsterdam. (2020). Raadscommissies - Gemeente Amsterdam. Retrieved December 21, 2020, from <https://www.amsterdam.nl/bestuur-organisatie/gemeenteraad/raadscommissies/>
- George, B., Baekgaard, M., Decramer, A., Audenaert, M., & Goeminne, S. (2020). Institutional isomorphism, negativity bias and performance information use by politicians: A survey experiment. *Public Administration*, 98(1), 14–28.
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. W.W. Norton & Company.
- Ghislieri, C., Molino, M., & Cortese, C. G. (2018). Work and organizational psychology looks at the Fourth Industrial Revolution: How to support workers and organizations? *Frontiers in Psychology*, 9.
- Gigerenzer, G. (1991). How to make Cognitive Illusions Disappear: Beyond "Heuristics and Biases." *European Review of Social Psychology*, 2(1), 83–115.
- Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1(1), 107–143.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review*, 103(4), 650–669.
- Gigerenzer, G., Hertwig, R., Hoffrage, U., & Sedlmeier, P. (2008). Cognitive Illusions Reconsidered. In *Handbook of Experimental Economics Results* (pp. 1018–1034).

- Gilad, S. (2019). Mixing Qualitative and Quantitative Methods in Pursuit of Richer Answers to Real-World Questions. *Public Performance and Management Review*, 44(5), 1075–1099.
- Gill, J. M., Chen, Y. X., Glutting, J. J., Diamond, J. J., & Lieberman, M. I. (2009). Impact of Decision Support in Electronic Medical Records on Lipid Management in Primary Care. *Population Health Management*, 12(5), 221–226.
- Gofen, A., Moseley, A., Thomann, E., & Weaver, R. K. (2021). Behavioural governance in the policy process: introduction to the special issue. *Journal of European Public Policy*, 28(5), 633–657.
- Goodhue, D., & Thompson, R. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19(2), 213–236.
- Gould, I. M., & Lawes, T. (2016). Antibiotic stewardship: prescribing social norms. *The Lancet*, 387, 1699–1701.
- Grady, K. E., Lemkau, J. P., Lee, N. R., & Caddell, C. (1997). Enhancing Mammography Referral in Primary Care. *Preventive Medicine*, 26(6), 791–800.
- Grant, A. M., & Hofmann, D. A. (2011). It's not all about me motivating hand hygiene among health care professionals by focusing on patients. *Psychological Science*, 22(12), 1494–1499.
- Greenberg, J., & Colquitt, J. (2005). *Handbook of organizational justice*. Lawrence Erlbaum Associates.
- Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: a movement in crisis? *British Medical Journal*, 348, g3725.
- Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral Public Administration: Combining Insights from Public Administration and Psychology. *Public Administration Review*, 77(1), 45–56.
- Grimshaw, J. M., Eccles, M. P., Lavis, J. N., Hill, S. J., & Squires, J. E. (2012). Knowledge translation of research findings. *Implementation Science*, 7(1), 50.
- Groeneveld, S., Tummers, L., Bronkhorst, B., Ashikali, T., & Van Thiel, S. (2015). Quantitative methods in public administration: Their use and development through time. *International Public Management Journal*, 18(1), 61–86.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293–323.
- Guest, G., Bunce, A., & Johnson, L. (1995). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability and a good number of journals in the. *Field Methods*, 18(1), 59–82.
- Guthrie, C., & Orr, D. (2006). Anchoring, Information, Expertise, and Negotiation: New Insights from Meta-Analysis. *Ohio State Journal on Dispute Resolution*, 21(3), 597–628.
- Hagger, M. S., & Luszczynska, A. (2014). Implementation Intention and Action Planning Interventions in Health Contexts: State of the Research and Proposals for the Way Forward. *Applied Psychology: Health and Well-Being*, 6(1), 1–47.
- Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., Clements, L., & Davies, S. C. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: A pragmatic national randomised controlled trial. *The Lancet*, 387(10029), 1743–1752.
- Hallsworth, M., Egan, M., Rutter, J., & Mccrae, J. (2018). Behavioural Government: Using behavioural science to improve how governments make decisions. The Behavioural Insights Team.
- Halpern, S. D., Ubel, P. A., & Asch, D. A. (2007). Harnessing the Power of Default Options to Improve Health Care. *The New England Journal of Medicine*, 356(27), 1340–1344.
- Han, X., Zhou, S., Fahoum, N., Wu, T., Gao, T., Shamay-Tsoory, S., Gelfand, M. J., Wu, X., & Han, S. (2021). Cognitive and neural bases of decision-making causing civilian casualties during intergroup conflict. *Nature Human Behaviour*, 5(9), 1214–1225.

References

- Hands, D. W. (2014). Normative ecological rationality: normative rationality in the fast-and-frugal-heuristics research program. *Journal of Economic Methodology*, 21(4), 396–410.
- Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the Manipulation of Choice. *European Journal of Risk Regulation*, 4(1), 3–28.
- Harrison, G. W., List, J. A., Burks, S., Camerer, C., Carpenter, J., Gerking, S., Isaac, R. M., Krueger, A., Mcmillan, J., Ortmann, A., Plott, C., Reiley, D., Rutström, E. E., & Wilcox, N. (2004). Field Experiments. *Journal of Economic Literature*, 42(4), 1009–1055.
- Harvard Law School. (2018). Harvard project will use behavioral insights to improve health care decisions and delivery - Harvard Law Today. Retrieved November 20, 2018, from <https://today.law.harvard.edu/harvard-project-will-use-behavioral-insights-improve-health-care-decisions-delivery/>
- Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The Evolution of Cognitive Bias. In *The Handbook of Evolutionary Psychology* (pp. 1–20). John Wiley & Sons, Inc.
- Hattke, F., Hensel, D., & Kalucza, J. (2020). Emotional Responses to Bureaucratic Red Tape. *Public Administration Review*, 80(1), 53–63.
- Helder, O. K., Weggelaar, A. M., Waarsenburg, D. C., Looman, C. W., van Goudoever, J. B., Brug, J., & Kornelisse, R. F. (2012). Computer screen saver hand hygiene information curbs a negative trend in hand hygiene behavior. *American Journal of Infection Control*, 40(10), 951–954.
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, 12(6), 973–986.
- Heukelom, F. (2012). A sense of mission: The alfred P. Sloan and Russell Sage foundations' behavioral economics program, 1984–1992. *Science in Context*, 25(2), 263–286.
- Hicks, L. S., Sequist, T. D., Ayanian, J. Z., Shaykevich, S., Fairchild, D. G., Orav, E. J., & Bates, D. W. (2008). Impact of Computerized Decision Support on Blood Pressure Management and Control: A Randomized Controlled Trial. *Journal of General Internal Medicine*, 23(4), 429–441.
- Hill, M., & Varone, F. (2021). *The public policy process*. Taylor & Francis Ltd.
- Hill, R. K. (2016). What an Algorithm Is. *Philosophy & Technology*, 29(1), 35–59.
- Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, 69(6), 1069–1086.
- Holland, J. H. (2014). *Complexity: A very short introduction*. OUP Oxford.
- Holm, J. M. (2017). Double standards? How historical and political aspiration levels guide managerial performance information use. *Public Administration*, 95(4), 1026–1040.
- Hong, S. W., Ching, T. Y., Fwng, J. P. M., & Seto, W. L. (1990). The employment of ward opinion leaders for continuing education in the hospital. *Medical Teacher*, 12(2), 209–217.
- Huis in 't Veld, N. M., Nagtegaal, R., & Noordegraaf, M. (2020). Duwtjes of druk? De percepties van zorgprofessionals aangaande 'nudging' in ziekenhuizen. *Bestuurskunde*, 29(4), 56–65.
- Hulgan, T., Rosenbloom, S. T., Hargrove, F., Talbert, D. A., Arbogast, P. G., Bansal, P., Miller, R. A., & Kernodle, D. S. (2004). Oral quinolones in hospitalized patients: an evaluation of a computerized decision support intervention. *Journal of Internal Medicine*, 256(4), 349–357.
- Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80(June 2019), 47–58.
- Hung, C.-S. S., Lin, J.-W. W., Hwang, J.-J. J., Tsai, R.-Y. Y., & Li, A.-T. T. (2008). Using paper chart based clinical reminders to improve guideline adherence to lipid management. *Journal of Evaluation in Clinical Practice*, 14(5), 861–866.
- Information Commissioner's Office. (2019). What should we do when we receive a request for information? ICO. Retrieved September 25, 2019, from <https://ico.org.uk/for-organisations/guide-to-freedom-of-information/>

- information/receiving-a-request/
- Ioannidis, J. P. A. (2018). Why most published research findings are false. In *Getting to Good: Research Integrity in the Biomedical Sciences* (pp. 2–8). Springer International Publishing.
- Ioannidis, J. P. A., Patsopoulos, N. A., & Rothstein, H. R. (2008). Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ*, 336(7658), 1413–1415.
- Irving, J. L. (1971). Groupthink. *Psychology Today*, 5(6), 43–46.
- Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: a meta-analysis of default effects. *Behavioural Public Policy*, 3(02), 159–186.
- Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371–377.
- Jilke, S., Olsen, A. L., Resh, W., & Siddiki, S. (2019). Microbrook, Mesobrook, Macrobrook. Perspectives on Public Management and Governance, 2(4), 245–253.
- Jilke, S., Petrovsky, N., Meuleman, B., & James, O. (2017). Measurement equivalence in replications of experiments: when and why it matters and guidance on how to determine equivalence. *Public Management Review*, 19(9), 1293–1310.
- Jilke, S., & Tummers, L. (2018). Which clients are deserving of help? A theoretical model and experimental test. *Journal of Public Administration Research and Theory*, 28(2), 226–238.
- John, P. (2018a). How far to nudge?: Assessing behavioural public policy. Edward Elgar Publishing.
- John, P. (2018b). Nudge: All tools are informational now. In *How Far to Nudge?* (pp. 55–68). Edward Elgar Publishing.
- John, P., & Blume, T. (2018). How best to nudge taxpayers? The impact of message simplification and descriptive social norms on payment rates in a central London local authority. *Journal of Behavioral Public Administration*, 1(1), 1–11.
- John, P., Smith, G., & Stoker, G. (2009). Nudge Nudge, Think Think: Two Strategies for Changing Civic Behaviour. *The Political Quarterly*, 80(3), 361–370.
- Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649), 1338–1339.
- Johnson, E. J., Shu, S. B., Dellaert, B. G. C., Fox, C., Goldstein, D. G., Häubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., Wansink, B., Weber, E. U., Johnson, E. J., Weber, E. U., Shu, S. B., Fox, C., Dellaert, B. G. C., Goldstein, D. G., Häubl, G., ... Wansink, B. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23(2), 487–504.
- Jones, B. D. (2017). Behavioral rationality as a foundation for public policy studies. *Cognitive Systems Research*, 43, 63–75.
- Jones, M. L. (2017). The right to a human in the loop: Political constructions of computer automation and personhood. *Social Studies of Science*, 47(2), 216–239.
- Jones, R., Pykett, J., & Whitehead, M. (2013). Psychological governance and behaviour change. *Policy and Politics*, 41(2), 159–182.
- Jousimaa, J., Mäkelä, M., Kunnamo, I., MacLennan, G., & Grimshaw, J. M. (2002). Primary care guidelines on consultation practices: the effectiveness of computerized versus paper-based versions. A cluster randomized controlled trial among newly qualified primary care physicians. *International Journal of Technology Assessment in Health Care*, 18(3), 586–596.
- Junghans, C. (2007). Effect of Patient-Specific Ratings vs Conventional Guidelines on Investigation Decisions in Angina. *Archives of Internal Medicine*, 167(2), 195–202.
- Kahan, N. R., Waitman, D.-A., & Vardy, D. A. (2009). Curtailing laboratory test ordering in a managed care setting through redesign of a computerized order form. *The American Journal of Managed Care*, 15(3), 173–176.

References

- Kahneman, D. (2003). A psychological perspective on economics. *American Economic Review*, 93(2), 162–168.
- Kahneman, D. (2011). Thinking fast and slow. Penguin Books Ltd.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), 193–206.
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). Noise. William Collins .
- Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press.
- Kahneman, D., & Tversky, A. (1996). On the Reality of Cognitive Illusions. *Psychological Review*, 103(3), 582–591.
- Kallinikos, J. (2005). The order of technology: Complexity and control in a connected world. *Information and Organization*, 15(3), 185–202.
- Kallinikos, J. (2009). Institutional complexity and functional simplification: the case of money claim online service in England and Wales. In *ICT and Innovation in the Public Sector* (pp. 174–210). Palgrave Macmillan UK.
- Karadimas, N. V., Papatzelou, K., & Loumos, V. G. (2007). Genetic Algorithms for Municipal Solid Waste Collection and Routing Optimization. In *Artificial Intelligence and Innovations 2007: from Theory to Applications* (pp. 223–231). Springer US.
- Kennedy, J. (1995). Debiasing the curse of knowledge in audit judgment. *Accounting Review*, 70(2), 249–273.
- Kent Pension Fund. (2020). About the Local Government Pension Scheme (LGPS). Kent County Council.
- Keren, G. (1990). Cognitive Aids and Debiasing Methods: CAN Cognitive Pills Cure Cognitive Ills? *Advances in Psychology*, 68, 523–552.
- Kiefe, C. I., Allison, J. J., Williams, O. D., Person, S. D., Weaver, M. T., & Weissman, N. W. (2001). Improving Quality Improvement Using Achievable Benchmarks For Physician Feedback. *Journal of the American Medical Association*, 285(22), 2871–2879.
- King, D., Greaves, F., Vlaev, I., & Darzi, A. (2013). Approaches based on behavioral economics could help nudge patients and providers toward lower health spending growth. *Health Affairs*, 32(4), 661–668.
- King, D., Vlaev, I., Everett-Thomas, R., Fitzpatrick, M., Darzi, A., & Birnbach, D. J. (2016). “Priming” Hand Hygiene Compliance in Clinical Environments. *Health Psychology*, 35(1), 96–101.
- Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *Conference on Human Factors in Computing Systems - Proceedings*, 2390–2395.
- Knies, E., Boselie, P., Gould-Williams, J., & Vandebaele, W. (2017). Strategic human resource management and public sector performance: context matters. *International Journal of Human Resource Management*, 1–13.
- Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, 110(3), 499.
- Kousgaard, M. B., Siersma, V., Reventlow, S., Ertmann, R., Felding, P., & Waldorff, F. B. (2013). The effectiveness of computer reminders for improving quality assessment for point-of-care testing in general practice—a randomized controlled trial. *Implementation Science*, 8(1), 47.
- Krause, G. A. (2006). Beyond the Norm. *Rationality and Society*, 18(2), 157–191.
- Kucher, N., Koo, S., Quiroz, R., Cooper, J. M., Paterno, M. D., Soukonnikov, B., & Goldhaber, S. Z. (2005). Electronic Alerts to Prevent Venous Thromboembolism among Hospitalized Patients. *New England Journal of Medicine*, 352(10), 969–977.
- Kullgren, J., Krupka, E., Schachter, A., & Linden, A. (2018). Precommitting to choose wisely about low-value services: a stepped wedge cluster randomised trial. 27(5), 355–364.

- Kwok, Y. L. A., Juergens, C. P., & McLaws, M. L. (2016). Automated hand hygiene auditing with and without an intervention. *American Journal of Infection Control*, 44(12), 1475–1480.
- Kwon, H. R., & Silva, E. A. (2019). Mapping the Landscape of Behavioral Theories: Systematic Literature Review: *Journal of Planning Literature*, 35(2), 161–179.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
- Landemore, H. (2017). Deliberative Democracy as Open, Not (Just) Representative Democracy. *Daedalus*, 146(3), 51–63.
- Larrick, R. P. (2004). Debiasing. In *Blackwell Handbook of Judgment and Decision Making* (pp. 316–338). Blackwell Publishing Ltd.
- Larsen, R. A., Evans, R. S., Burke, J. P., Pestotnik, S. L., Gardner, R. M., & Classen, D. C. (1989). Improved perioperative antibiotic use and reduced surgical wound infections through use of computer decision analysis. *Infection Control & Hospital Epidemiology*, 10(7), 316–320.
- Lasswell, H. (1936). Politics: Who gets what, when, how. Pickle Partners Publishing.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668.
- Leventhal, G. S. (1980). What Should Be Done with Equity Theory? In *Social Exchange* (pp. 27–55). Springer US.
- Lewis, A. (2008). *The Cambridge handbook of psychology and economic behaviour*. Cambridge University Press.
- Liao, J. M., Fleisher, L. A., Navathe, A. S., Meeker, D., McDonald, R., Waldo, S. W., Ferraro, P. J., & Baron, R. J. (2016). Increasing the Value of Social Comparisons of Physician Performance Using Norms. *Journal of the American Medical Association*, 316(11), 562–570.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *PLoS Med*, 6(7), e1000100.
- Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D. (2017). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 25(1), 322–349.
- Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science*, 4(4), 390–398.
- Lind, E., & Tyler, T. (1988). *The Social Psychology of Procedural Justice*. Springer Science & Business Media.
- Lind, E., & Tyler, T. (1988). The social psychology of procedural justice. Springer Science & Business Media.
- Linos, E., Reinhard, J., & Ruda, S. (2017). Levelling the playing field in police recruitment: Evidence from a field experiment on test performance. *Public Administration*, 95(4), 943–956.
- Lipsky, M. (2010). *Street-level bureaucracy*. Russell Sage Foundation.
- Liu, X., Stoutenborough, J., & Vedlitz, A. (2017). Bureaucratic expertise, overconfidence, and policy choice. *Governance*, 30(4), 705–725.
- Lo, H. G., Matheny, M. E., & Seger, D. L. (2009). Impact of non-interruptive medication laboratory monitoring alerts in ambulatory care. *Journal of the American Medical Informatics Association*, 16(1), 66–71.
- Lodge, M., & Wegrich, K. (2016). The Rationality Paradox of Nudge: Rational Tools of Government in a World of Bounded Rationality. *Law & Policy*, 38(3), 250–267.
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.

References

- Loibl, C., Sunstein, C. R., Rauber, J., & Reisch, L. A. (2018). Which Europeans Like Nudges? Approval and Controversy in Four European Countries. *Journal of Consumer Affairs*, 52(3), 655–688.
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231–1243.
- Luangasanatip, N., Hongsuwan, M., Limmathurotsakul, D., Lubell, Y., Lee, A. S., Harbarth, S., Day, N. P. J., Graves, N., & Cooper, B. S. (2015). Comparative efficacy of interventions to promote hand hygiene in hospital: systematic review and network meta-analysis. *BMJ*, 351, h3728.
- Lucke, J. (2013). Context Is All Important in Investigating Attitudes: Acceptability Depends on the Nature of the Nudge, Who Nudges, and Who Is Nudged. *American Journal of Bioethics*, 13(6), 24–25.
- Ludolph, R., & Schulz, P. J. (2018). Debiasing Health-Related Judgments and Decision Making: A Systematic Review. *Medical Decision Making*, 38(1), 3–13.
- Lupo, G. (2019). Assessing e-justice smartness: A new framework for e-justice evaluation through public values. In *Public Administration and Information Technology* (Vol. 35, pp. 77–113). Springer.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70(3, Pt.1), 151–159.
- Mafi, J. N., & Parchman, M. (2018). Low-value care: an intractable global problem with no quick fix. *BMJ Quality & Safety*, 27(5), 333–336.
- Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M. J. P., van Saase, J. L. C. M., van Guldener, C., & Schmidt, H. G. (2010). Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents. *Journal of the American Medical Association*, 304(11), 1198.
- Manfreda, K. L., Batagelj, Z., & Vehovar, V. (2006). Design of Web Survey Questionnaires: Three Basic Experiments. *Journal of Computer-Mediated Communication*, 7(3), JCMC731.
- March, J. G., & Olsen, J. P. (1983). The New Institutionalism: Organizational Factors in Political Life. *American Political Science Review*, 78(03), 734–749.
- March, J. G., & Olsen, J. P. (2019). The Logic of Appropriateness. In R. E. Goodin (Ed.), *The Oxford Handbook of Political Science*.
- Marewski, J. N., & Gigerenzer, G. (2012). Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience*, 14(1), 77–89.
- Margetts, H. Z. (2011). Experiments for Public Management Research. *Public Management Review*, 13(2), 189–208.
- Marin-Garcia, J. A., & Tomas, J. M. (2016). Deconstructing AMO framework: A systematic review Intangible Deconstructing AMO framework: A systematic review. *Intangible Capital*, 12(4), 1040–1087.
- Martens, J. D., van der Weijden, T., Severens, J. L., de Clercq, P. A., de Bruijn, D. P., Kester, A. D. M., & Winkens, R. A. G. (2007). The effect of computer reminders on GPs' prescribing behaviour: A cluster-randomised trial. *International Journal of Medical Informatics*, 76(3), S403–S416.
- Masic, I., Miokovic, M., & Muhammedagic, B. (2008). Evidence Based Medicine – New Approaches and Challenges. *Acta Informatica Medica*, 16(4), 219.
- Maynard-Moody, S., & Musheno, M. (2000). State Agent or Citizen Agent: Two Narratives of Discretion. *Journal of Public Administration Research and Theory*, 10(2), 329–358.
- Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. R. (2013). Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice. *Criminology*, 51(1), 33–63.
- McGowan, L. J., Powell, R., & French, D. P. (2020). How can use of the Theoretical Domains Framework be optimized in qualitative research? A rapid systematic review. *British Journal of Health Psychology*, 25(3), 677–694.

- Meder, B., Fleischhut, N., & Osman, M. (2018). Beyond the confines of choice architecture: A critical analysis. *Journal of Economic Psychology*, 68, 36–44.
- Meeker, D., Knight, T. K., Friedberg, M. W., Linder, J. A., Goldstein, N. J., Fox, C. R., Rothfeld, A., Diaz, G., & Doctor, J. N. (2014). Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. *JAMA Internal Medicine*, 174(3), 425–431.
- Meeker, D., Linder, J. A., Fox, C. R., Friedberg, M. W., Persell, S. D., Goldstein, N. J., Knight, T. K., Hay, J. W., & Doctor, J. N. (2016). Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices. *JAMA*, 315(6), 562.
- Meier, K. J., Winter, S. C., O'Toole, L. J., Favero, N., & Andersen, S. C. (2015). The validity of subjective performance measures: school principles in Texas and Denmark. *Public Administration*, 93(4), 1084–1101.
- Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration*, 42(12), 1031–1039.
- Melnick, E. R., Genes, N. G., Chawla, N. K., Akerman, M., Baumlin, K. M., & Jagoda, A. (2010). Knowledge translation of the American college of emergency physicians' Clinical Policy on Syncope using computerized clinical decision support. *International Journal of Emergency Medicine*, 3(2), 97–104.
- Messing, J. (2015). Improving handover from intensive care to ward medical teams with simple changes to paperwork. *BMJ Open Quality*, 4(1), u206467-w2913.
- Michie, S., Atkins, L., & West, R. (2014). *The Behaviour Change Wheel Book - A Guide To Designing Interventions*. Silverback Publishing.
- Michie, S., Johnston, M., Abraham, C., Lawton, R., Parker, D., & Walker, A. (2005). Making psychological theory useful for implementing evidence based practice: a consensus approach. *Quality & Safety in Health Care*, 14(1), 26–33.
- Michie, S., & Lester, K. (2005). Words matter: increasing the implementation of clinical guidelines. *Quality Safety in Health Care*, 14(5), 367–370.
- Michie, S., Thomas, J., Johnston, M., Aonghusa, P. Mac, Shawe-Taylor, J., Kelly, M. P., Deleris, L. A., Finnerty, A. N., Marques, M. M., Norris, E., O'Mara-Eves, A., & West, R. (2017). The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science*, 12(1), 1–12.
- Michie, S., van Stralen, M. M., West, R., Grimshaw, J., Shirran, L., Thomas, R., Mowatt, G., Fraser, C., Bero, L., & Coleman, T. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1), 42.
- Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How Can Decision Making Be Improved? *Perspectives on Psychological Science*, 4(4), 379–383.
- Mills, S. (2020). Personalized nudging. *Behavioural Public Policy*, 1–10.
- Mitchell, M. (2009). *Complexity : A Guided Tour*. Oxford University Press, USA.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology*, 63(8), e1–e37.
- Moore, M. (1995). *Creating public value: Strategic management in government*. Harvard university press.
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 129–140.
- Moscrop, D. R., & Warren, M. E. (2016). When is Deliberation Democratic? *Journal of Public Deliberation*, 12(2).

References

- Moseley, A., & Thomann, E. (2021). A behavioural model of heuristics and biases in frontline policy implementation. *Policy and Politics*, 49(1), 49–67.
- Moynihan, D. (2018). A great schism approaching? Towards a micro and macro public administration. *Journal of Behavioral Public Administration*, 1(1), 1–8.
- Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions. *Journal of Public Administration Research and Theory*, 25(1), 43–69.
- Moynihan, D. P., & Lavertu, S. (2012). Cognitive Biases in Governing: Technology Preferences in Election Administration. *Public Administration Review*, 72(1), 68–77.
- Moynihan, D. P., Nielsen, P. A., & Kroll, A. (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. In O. James, S. R. Jilke, & G. G. van Ryzin (Eds.), *Experiments in Public Management Research*. Cambridge University Press.
- Münscher, R., Vetter, M., & Scheuerle, T. (2016). A Review and Taxonomy of Choice Architecture Techniques. *Journal of Behavioral Decision Making*, 29(5), 511–524.
- Murtaugh, C. M., Pezzin, L. E., McDonald, M. V., Feldman, P. H., & Peng, T. R. (2005). Just-in-Time Evidence-Based E-mail "Reminders" in Home Health Care: Impact on Nurse Practices. *Health Services Research*, 40(3), 849–864.
- Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility. *Personality and Social Psychology Bulletin*, 26(9), 1142–1150.
- Nagtegaal, R., Tummers, L., Noordegraaf, M., & Bekkers, V. (2019). Nudging Healthcare Professionals towards Evidence-Based Medicine: A Systematic Scoping Review. *Journal of Behavioral Public Administration*, 2(2).
- Nederlandse Vereniging voor Raadsleden. (2020a). Griffier. Retrieved December 21, 2020, from <https://www.raadsleden.nl/het-raadslid/wat-doet-de-gemeenteraad/griffier>
- Nederlandse Vereniging voor Raadsleden. (2020b). Wat doet de gemeenteraad? Retrieved April 30, 2020, from <https://www.raadsleden.nl/het-raadslid/wat-doet-de-gemeenteraad>
- Nevo, I., Fitzpatrick, M., Thomas, R.-E., Gluck, P. A., Lenchus, J. D., Arheart, K. L., & Birnbach, D. J. (2010). The Efficacy of Visual Cues to Improve Hand Hygiene Compliance. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 5(6), 325–331.
- Nicholson-Crotty, S., Nicholson-Crotty, J., & Webeck, S. (2019). Are public managers more risk averse? Framing effects and status quo bias across the sectors. *Journal of Behavioral Public Administration*, 2(1).
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175–220.
- Niehaves, B., & Plattfaut, R. (2014). Internet adoption by the elderly: Employing IS technology acceptance theories for understanding the age-related digital divide. *European Journal of Information Systems*, 23(6), 708–726.
- Nilsen, P., & Bernhardsson, S. (2019). Context matters in implementation science: A scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Services Research*, 19(1), 1–21.
- Nitzl, C., Sicilia, M., & Steccolini, I. (2019). Exploring the links between different performance information uses, NPM cultural orientation, and organizational performance in the public sector. *Public Management Review*, 21(5), 686–710.
- Noordegraaf, M. (2015). *Public Management. Performance, Professionalism and Politics*. Palgrave.
- Noordegraaf, M., & Abma, T. (2003). Management by Measurement? Public Management Practices Amidst Ambiguity. *Public Administration*, 81(4), 853–871.

- Nosek, B. A., & Lakens, D. (2014). Registered reports : a method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141.
- O'Connor, C., Adhikari, N. K. J. J., DeCaire, K., & Friedrich, J. O. (2009). Medical admission order sets to improve deep vein thrombosis prophylaxis rates and other outcomes. *Journal of Hospital Medicine*, 4(2), 81–89.
- O'Neill, C. (2016). Weapons of math destruction : how big data increases inequality and threatens democracy. Crown Publishing Group.
- OECD. (2012a). Human Resources Management Country Profiles: ITALY. In *Human Resources Management Country Profiles*.
- OECD. (2012b). Human Resources Management Country Profiles: UNITED KINGDOM. *Human Resources Management Country Profiles*.
- Olejniczak, K., Śliwowski, P., & Leeuw, F. (2020). Comparing Behavioral Assumptions of Policy Tools: Framework for Policy Designers. *Journal of Comparative Policy Analysis: Research and Practice*, 22(6), 498–520.
- Oliver, A. (2013). From Nudging to Budging: Using Behavioural Economics to Inform Public Sector Policy. *Journal of Social Policy*, 42(4), 1–16.
- Olson, J., Hollenbeck, C., Donaldson, K., Abendroth, T., & Castellani, W. (2015). Default settings of computerized physician order entry system order sets drive ordering habits. *Journal of Pathology Informatics*, 6(1), 16.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170359.
- Ottaviani, M., & Sørensen, P. (2001). Information aggregation in debate: Who should speak first? *Journal of Public Economics*, 81(3), 393–421.
- Overheid in Nederland. (2018). Rapportage kenmerken politieke ambtsdragers.
- Page, S. B., Stone, M. M., Bryson, J. M., & Crosby, B. C. (2015). Public value creation by cross-sector collaborations: A framework and challenges of assessment. *Public Administration*, 93(3), 715–732.
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.
- Parkin, J. (2011). Adaptable Due Process. *University of Pennsylvania Law Review*, 160(20), 1309–1378.
- Patel, M. S., Volpp, K. G., & Asch, D. A. (2018). Nudge Units to Improve the Delivery of Health Care. *New England Journal of Medicine*, 378(3), 214–216.
- Patey, A. M., Islam, R., Francis, J., Bryson, G. L., Grimshaw, J. M., Driedger, M., Eccles, M., Godin, G., Hux, J., Johnston, M., Légaré, F., Lemire, L., Pomey, M. P., & Sales, A. (2012). Anesthesiologists' and surgeons' perceptions about routine pre-operative testing in low-risk patients: Application of the Theoretical Domains Framework (TDF) to identify factors that influence physicians' decisions to order pre-operative tests. *Implementation Science*, 7(1), 1–13.
- Patil, P., Peng, R. D., & Leek, J. T. (2016). What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science*, 11(4), 539–544.
- Patterson, R. (1998). A computerized reminder for prophylaxis of deep vein thrombosis in surgical patients. *Proceedings of the AMIA Symposium*, 573–576.
- Pedersen, M. J., & Stritch, J. M. (2018a). Internal Management and Perceived Managerial Trustworthiness: Evidence From a Survey Experiment. *The American Review of Public Administration*, 48(1), 67–81.

References

- Pedersen, M. J., & Stritch, J. M. (2018b). RNICE Model: Evaluating the Contribution of Replication Studies in Public Administration and Management Research. *Public Administration Review*, 78(4), 606–612.
- Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., & Frik, A. (2020). Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior*, 109, 106347.
- Perry, J. L., & Vandenabeele, W. (2015). Public Service Motivation Research: Achievements, Challenges, and Future Directions. *Public Administration Review*, 75(5), 692–699.
- Peters, K., Boogaard, G., van den Berg, B., & van Kalken, L. (2020). Derde rapportage, eindrapport evaluatiecommissie tijdelijke wet digitale beraadslaging en besluitvorming | Kennisbank Openbaar Bestuur.
- Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare*, 13(3), 141–146.
- Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research Synthesis Methods*, 5(4), 371–385.
- Phillips, C. J., Marshal, A. P., Chaves, N. J., Jankelowitz, S. K., Lin, I. B., Loy, C. T., Rees, G., Sakzewski, L., Thomas, S., To, T. P., Wilkinson, S. A., & Michie, S. (2015). Experiences of using the Theoretical Domains framework across diverse clinical environments: A qualitative study. *Journal of Multidisciplinary Healthcare*, 8, 139–146.
- Poley, M. J., Edelenbos, K. I., Mosseveld, M., Van Wijk, M. A. M., De Bakker, D. H., Van Der Lei, J., & Rutten-Van Möhlen, M. P. M. H. (2007). Cost consequences of implementing an electronic decision support system for ordering laboratory tests in primary care: Evidence from a controlled prospective study in the Netherlands. *Clinical Chemistry*, 53(2), 213–219.
- Pollitt, C. (2013). Context in public policy and management: The missing link? (C. Pollitt (ed.)). Edward Elgar Publishing.
- Pollitt, C., & Bouckaert, G. (2011). Public management reform: a comparative analysis. Oxford University Press.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65–76.
- Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14.
- Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556.
- Raub, M. (2018). Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. *Arkansas Law Review*, 71.
- Reddick, C. G. (2005). Citizen interaction with e-government: From the streets to servers? *Government Information Quarterly*, 22(1), 38–57.
- Repubblica.it. (2016). Ecco il Freedom of Information Act, la versione definitiva. Retrieved May, 05, 2020, from https://www.repubblica.it/tecnologia/sicurezza/2016/05/19/news/ecco_il_freedom_of_information_act_la_versione_definitiva-140129607.
- Richter, T., Fishbain, B., Markus, A., Richter-Levin, G., & Okon-Singer, H. (2020). Using machine learning-based analysis for behavioral differentiation between anxiety and depression. *Scientific Reports*, 10(1), 1–12.
- Roberts, A. (2020). Bridging Levels of Public Administration: How Macro Shapes Meso and Micro. *Administration & Society*, 52(4), 631–656.

- Rood, E., Bosman, R. J., Van Der Spoel, J. I., Taylor, P., & Zandstra, D. F. (2005). Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation. *Journal of the American Medical Informatics Association*, 12(2), 172–180.
- Roukema, J., Steyerberg, E. W., van der Lei, J., & Moll, H. A. (2008). Randomized trial of a clinical decision support system: impact on the management of children with fever without apparent source. *Journal of the American Medical Informatics Association*, 15(1), 107–113.
- Rubin, E. V. (2007). The Role of Procedural Justice in Public Personnel Management: Empirical Results from the Department of Defense. *Journal of Public Administration Research and Theory*, 19(1), 125–143.
- Sackett, D. J., Rosenberg, W. M., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, 71–72.
- Sadule-Rios, N., & Aguilera, G. (2017). Nurses' perceptions of reasons for persistent low rates in hand hygiene compliance. *Intensive and Critical Care Nursing*, 42, 17–21.
- Samuels, R., Stich, S., & Bishop, M. (2002). Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear. Oxford University Press.
- Sanders, M., Snijders, V., & Hallsworth, M. (2018). Behavioural science and policy: where are we now and where are we going? *Behavioural Public Policy*, 2(2), 144–167.
- Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 497–502.
- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599.
- Schafer, B. A., & Schafer, J. K. (2018). Client Likeability in Auditor Fraud Risk Judgments: The Mitigating Influence of Task Experience, the Review Process, and a "Consider the Opposite" Strategy. *Current Issues in Auditing*, 12(1), P11–P16.
- Scheerder, A., Deursen, A. van, & Dijk, J. van. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and Informatics*, 34(8), 1607–1624.
- Schillemans, T., & Giesen, I. (2020). De kwaliteit van (juridische) oordelen. *Nederlands Juristenblad*, 95(3), 184–193.
- Schwann, N. M., Bretz, K. A., Eid, S., Burger, T., Fry, D., Ackler, F., Evans, P., Romancheck, D., Beck, M., Ardire, A. J., Lukens, H., & McLoughlin, T. M. (2011). Point-of-care electronic prompts: An effective means of increasing compliance, demonstrating quality, and improving outcome. *Anesthesia and Analgesia*, 113(4), 869–876.
- Secunda, P. M. (2012). Cognitive Illiberalism and Institutional Debiasing Strategies. *San Diego Law Review*, 49(2), 373–414.
- Sela, A. (2019). E-Nudging Justice: The Role of Digital Choice Architecture in Online Courts. *Journal of Dispute Resolution*, 2019, 127–164.
- Sequist, T. D., Gandhi, T. K., Karson, A. S., Fiskio, J. M., Bugbee, D., Sperling, M., Cook, E. F., Fairchild, D. G., & Bates, D. W. (2005). A randomized trial of electronic clinical reminders to improve quality of care for diabetes and coronary artery disease. *Journal of the American Medical Informatics Association*, 12(4), 431–437.
- Shank, D. B. (2016). Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk. *American Sociologist*, 47(1), 47–55.
- Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, 85(1), 140–156.

References

- Sheeran, P. (2002). Intention—Behavior Relations: A Conceptual and Empirical Review. *European Review of Social Psychology*, 12(1), 1–36.
- Sheffer, L., Loewen, P. J., Soroka, S., Walgrave, S., & Sheaffer, T. (2018). Nonrepresentative representatives: An experimental study of the decision making of elected politicians. *American Political Science Review*, 112(2), 302–321.
- Simon, H. A. (1997). *Administrative Behavior* (Fourth Edition). The Free Press.
- Smith, B. W., & Slack, M. B. (2015). The effect of cognitive debiasing training among family medicine residents. *Diagnosis*, 2(2), 117–121.
- Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A User's Guide to Debiasing. In *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 924–951). John Wiley & Sons, Ltd.
- Soman, D., & Yeung, C. (2021). The Behaviourally Informed Organization. In *The Behaviourally Informed Organization*. University of Toronto Press.
- Sousa Lourenço, J., Emanuele Ciriolo, E., Rafael Almeida, S., & Xavier Troussard, X. (2016). *Behavioural Insights Applied to Policy*.
- Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, 5(2), 199–222.
- Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–726.
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring Political Deliberation: A Discourse Quality Index. *Comparative European Politics*, 1(1), 21–48.
- Stern, J. M., & Simes, R. J. (1997). Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ (Clinical Research Edition)*, 315(7109), 640–645.
- Stone, D. (1997). *Policy paradox: The art of political decision making*. Ww Norton & Co.
- Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73(3), 437–446.
- Strandberg, K., & Grönlund, K. (2018). Online Deliberation. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. E. Warren (Eds.), *The Oxford Handbook of Deliberative Democracy*.
- Straßheim, H., Korinek, R., Strassheim, H., Jung, A., & Korinek, R.-Iea. (2015). Reframing expertise: The rise of behavioural insights and interventions in public policy . In *Moments of valuation. Exploring sites of dissonance* (pp. 249–268).
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Sage publications.
- Strohmeier, S., & Piazza, F. (2013). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, 40(7), 2410–2420.
- Strom, B., Schinnar, R., & Bilker, W. (2010). Randomized clinical trial of a customized electronic alert requiring an affirmative response compared to a control group receiving a commercial passive CPOE. *Journal of the American Medical Informatics Association*, 17(4), 411–415.
- Sunshine, J., & Tyler, T. R. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law Society Review*, 37(3), 513–548.
- Sunstein, C. R. (2012). Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych. *SSRN Electronic Journal*.
- Sunstein, C. R. (2014). Nudging: a very short guide. *Journal of Consumer Policy*, 37(4), 583–588.
- Sunstein, C. R. (2016). People Prefer System 2 Nudges (Kind of). *Duke Law Journal*, 66(1), 121–168.
- Sunstein, C. R. (2017). Nudges that fail. *Behavioural Public Policy*, 1(1), 4–25.

- Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4), 1159-1202.
- Sunstein, C., & Thaler, R. (2008). *Nudge. The Politics of Libertarian Paternalism*. New Haven.
- Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal*, 11(2), 63-75.
- Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2017). A Systematic Scoping Review of the Choice Architecture Movement: Toward Understanding When and Why Nudges Work. *Journal of Behavioral Decision Making*, 31(3), 355-366.
- Tannenbaum, D., Doctor, J. N., Persell, S. D., Friedberg, M. W., Meeker, D., Friesema, E. M., Goldstein, N. J., Linder, J. A., & Fox, C. R. (2015). Nudging Physician Prescription Decisions by Partitioning the Order Set: Results of a Vignette-Based Study. *Journal of General Internal Medicine*, 30(3), 298-304.
- Tannenbaum, D., Fox, C. R., & Rogers T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, 1(7), 1-7.
- Thaler, R. (2018). Nudge, not sludge. *Science*, 361(6401), 431.
- Thaler, R. H., & Benartzi, S. (2004). Save More TomorrowTM: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, 112(S1), S164-S187.
- Thaler, R. H., & Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press.
- The Cochrane Collaboration. (2018). Glossary | Cochrane Community. Retrieved May, 01, 2018, from <http://community.cochrane.org/glossary>
- The economist. (2008). Nudge, nudge | Free exchange | The Economist. Retrieved March 25, 2021, from <https://www.economist.com/free-exchange/2008/08/13/nudge-nudge>
- The Joanna Briggs Institute. (2015). The Joanna Briggs Institute Reviewers' Manual 2015 Methodology for JBI Scoping Reviews.
- The New York Times. (2018). Firm Led by Google Veterans Uses A.I. to 'Nudge' Workers Toward Happiness. Retrieved October 7, 2019, from <https://www.nytimes.com/2018/12/31/technology/human-resources-artificial-intelligence-humu.html>
- Thibodeau, P. H., & Boroditsky, L. (2015). Measuring effects of metaphor in a dynamic opinion landscape. *PLoS ONE*, 10(7), e0133939.
- Thomas, M., & Morwitz, V. (2005). Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition. *Journal of Consumer Research*, 32(1), 54-64.
- Tierney, W. M., Overhage, J. M., Murray, M. D., Harris, L. E., Zhou, X.-H., Eckert, G. J., Smith, F. E., Nienaber, N., McDonald, C. J., & Wolinsky, F. D. (2003). Effects of computerized guidelines for managing heart disease in primary care. *Journal of General Internal Medicine*, 18(12), 967-976.
- Tierney, W. M., Overhage, J. M., Murray, M. D., Harris, L. E., Zhou, X.-H., Eckert, G. J., Smith, F. E., Nienaber, N., McDonald, C. J., & Wolinsky, F. D. (2005). Can Computer-Generated Evidence-Based Care Suggestions Enhance Evidence-Based Management of Asthma and Chronic Obstructive Pulmonary Disease? A Randomized, Controlled Trial. *Health Services Research*, 40(2), 477-498.
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garrity, C., ... Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467.
- Tummers, L. (2011). Explaining the willingness of public professionals to implement new policies: A policy alienation framework. *International Review of Administrative Sciences*.
- Tummers, L. (2017). The Relationship Between Coping and Job Performance. *Journal of Public Administration Research and Theory*, 27(1), 150-162.

References

- Tummers, L. (2019). Public Policy and Behavior Change. *Public Administration Review*, 79(6), 925–930.
- Tummers, L. (2020). Behavioral Public Administration. *Oxford Research Encyclopedia of Politics*.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Twizeyimana, J. D., & Andersson, A. (2019a). The public value of E-Government – A literature review. In *Government Information Quarterly* (Vol. 36, Issue 2, pp. 167–178). Elsevier Ltd.
- Twizeyimana, J. D., & Andersson, A. (2019b). The public value of E-Government – A literature review. *Government Information Quarterly*, 36(2), 167–178.
- Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel Psychology*, 65(3), 597–660.
- Ünveren, B., & Baycar, K. (2019). Historical evidence for anchoring bias: The 1875 cadastral survey in Istanbul. *Journal of Economic Psychology*, 73, 1–14.
- Vaccaro, M., & Waldo, J. (2019). The effects of mixing machine learning and human judgment. *Communications of the ACM*, 62(11), 104–110.
- van Deursen, A. J. A. M., & Helsper, E. J. (2015). The Third-Level Digital Divide: Who Benefits Most from Being Online? (pp. 29–52). Emerald Group Publishing Limited.
- Van Deursen, A. J., & Van Dijk, J. A. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society*, 21(2), 354–375.
- Van Wyk, J. T., Van Wijk, M. A. M., Sturkenboom, M. C. J. M., Mosseveld, M., Moorman, P. W., & Van Der Lei, J. (2008). Electronic alerts versus on-demand decision support to improve dyslipidemia treatment: A cluster randomized controlled trial. *Circulation*, 117(3), 371–378.
- van Zoonen, L. (2016). Privacy concerns in smart cities. *Government Information Quarterly*, 33(3), 472–480.
- Vandenabeele, W. (2007). Toward a public administration theory of public service motivation. *Public Management Review*, 9(4), 545–556.
- Vaughn, V. M., & Linder, J. A. (2018). Thoughtless design of the electronic health record drives overuse, but purposeful design can nudge improved patient care. *BMJ Quality & Safety*, 27(8), 583–586.
- Veale, M., & Brass, I. (2019). Administration by Algorithm. In K. Yeung & M. Lodge (Eds.), *Algorithmic Regulation*. Oxford University Press.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly: Management Information Systems*, 27(3), 425–478.
- Verbiest, M. E., Presseau, J., Chavannes, N. H., Scharloo, M., Kaptein, A. A. A. A., Assendelft, W. J. J., & Crone, M. R. (2014). Use of action planning to increase provision of smoking cessation care by general practitioners: role of plan specificity and enactment. *Implementation Science*, 9(1), 180.
- Villegas, S., Lloyd, R. A., Tritt, A., & Vengrouskie, E. F. (2019). Human Resources as Ethical Gatekeepers: Hiring Ethics and Employee Selection. *Journal of Leadership, Accountability and Ethics*, 16(2), 80–88.
- Vis, B. (2019). Heuristics and Political Elites' Judgment and Decision-Making. *Political Studies Review*, 17(1), 41–52.
- Viscusi, W. K., & Gayer, T. (2015). Behavioral public choice: the behavioural paradox of government policy. *Harvard Journal of Law & Public Policy*, 38(3), 973–1007.

- Visintin, E. P., Bonvin, J. M., Varone, F., Butera, F., Lovey, M., & Rosenstein, E. (2021). Can street-level bureaucrats be nudged to increase effectiveness in welfare policy? *Policy and Politics*, 49(1), 121–139.
- VonVille, H. (2018). Excel Workbooks & User Guides for Systematic Reviews. Retrieved September 4, 2018, from <https://showcase.dropbox.com/s/Excel-Workbooks-User-Guides-for-Systematic-Reviews-Kf4pYVTvFqSjdZR2Spzlu>
- Walker, R. M., James, O., & Brewer, G. A. (2017). Replication, experiments and knowledge in public management research. *Public Management Review*, 19(9), 1221–1234.
- Weaver, R. K. (2015). Getting People to Behave: Research Lessons for Policy Makers. *Public Administration Review*, 75(6), 806–816.
- Weber, M. (2015). Bureaucracy. In J. Shafritz, J. Ott, & Y. Jang (Eds.), *Classics of organization theory* (Sixth edition, pp. 73–79).
- Weinmann, M., Schneider, C., & Brocke, J. vom. (2016). Digital Nudging. *Business and Information Systems Engineering*, 58(6), 433–436.
- Weir CJ, Lees KR, MacWalter RS., & et al. (2013). Cluster-randomized, controlled trial of computer-based decision support for selecting long-term anti-thrombotic therapy after acute ischaemic stroke. *QJM: An International Journal of Medicine*, 96(2), 143–153.
- Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior*, 101, 197–209.
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical Power and Optimal Design in Experiments in Which Samples of Participants Respond to Samples of Stimuli. *Journal of Experimental Psychology*, 143(5), 2020–2045.
- Wildavsky, A. (1973). If planning is everything, maybe it's nothing. *Policy Sciences*, 4(2), 127–153.
- Williamson, J. W., German, P. S., Weiss, R., Skinner, E. A., & Bowes, F. (1989). Health science information management and continuing education of physicians. A survey of U.S. primary care practitioners and their opinion leaders. *Annals of Internal Medicine*, 110(2), 151–160.
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116(1), 117–142.
- Woolf, S., Grol, R., Hutchinson, A., & Eccles, M. (1999). Potential benefits, limitations, and harms of clinical guidelines. *BMJ*, 318(7138), 527–530.
- Worthy, B. (2010). More open but not more trusted? The effect of the freedom of information act 2000 on the United Kingdom central government. *Governance*, 23(4), 561–582.
- WRR. (2017). Weten is nog geen doen. Een realistisch perspectief op redzaamheid | Rapport | WRR.
- Xia, J., Curtin, K. M., Li, W., & Zhao, Y. (2015). A New Model for a Carpool Matching Service. *PloS One*, 10(6), e0129257.
- Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118–136.
- Zarsky, T. (2016). The Trouble with Algorithmic Decisions. *Science, Technology, & Human Values*, 41(1), 118–132.
- Zingg, W., Castro-Sanchez, E., Secci, F. V., Edwards, R., Drumright, L. N., Sevdalis, N., & Holmes, A. H. (2016). Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). *Public Health*, 133, 19–37.
- Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated discretion. In P. Hupe & T. Evans (Eds.), *Palgrave Handbook on discretion: The quest for controlled freedom*. Palgrave.
- Zurovac, D., Sudoi, R. K., Akhwale, W. S., Ndiritu, M., Hamer, D. H., Rowe, A. K., & Snow, R. W. (2011). The effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. *The Lancet*, 378(9793), 795–803.

APPENDICES

APPENDIX A - SEARCH STRATEGY FOR CHAPTER 2

Specific search

((nudge or nudging "choice architecture" or "behavioural economics" or "behavioral economics") and (health care or healthcare or medical) and (practitioners or doctors or nurses or clinicians or surgeons) and (guidelines or "evidence based medicine")).af. and (experiment* or trial or intervention).ab. – Ovid Medline, PsychInfo

Broad search

("choice architect*" OR nudg* OR behavio*) AND (health care OR healthcare OR medic*)
AND (experiment* OR trial OR intervention) AND (practitioners OR doctors OR nurses OR
clinicians OR surgeons) AND (guidelines OR "evidence based medicine") [all] – PubMed

APPENDIX B – PRISMA STATEMENT FOR CHAPTER 2

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	29
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	30
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	31-33
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	32
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	35
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	34-35
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	35 and appendix A
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix A
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	36
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	36
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Appendix A
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	33
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I ²) for each meta-analysis.	N/A

A

Appendices

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	33, 40
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or sub-group analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	36
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Online Appendix*
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Online Appendix*
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	40
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or sub-group analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	42-45
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	45
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	46
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	N/A

Note. The Online Appendix is accessible at <https://dataverse.harvard.edu/dataverse/JBPA>

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed.1000097
For more information, visit: www.prisma-statement.org.

APPENDIX C. CONDITIONS FOR EXPERIMENTS FOR CHAPTER 3

Experiment 1: Bellé, Cantarelli, and Belardinelli (2018).

You are the senior manager of the Public Relations Office in a medium-sized municipality. You have to decide the maximum number of days by which your subordinates have to reply to citizens' inquiries sent via emails. Consider whether the maximum number of days to reply to citizens' emails must be higher or lower than 2[90] working days.

Consider-the-opposite.

To make this decision, please first list two reasons why 2[90] days might be too short[long] to respond to citizens' emails.

- Reason 1 _____
- Reason 2 _____

Now, indicate the maximum number of days to reply to citizens' emails below.

Experiment 2: Bellé, Cantarelli, and Belardinelli (2017).

Imagine that you have to assess this year's performance of a subordinate of yours. During this year, your subordinate met the majority of goals, had good interpersonal skills with her colleagues, and showed moderate creativity in proposing new ideas for the improvement of the services.

The previous year, you assigned your subordinate a performance rating of 51/100[91/100]. You have to decide whether to assign a rating lower or higher than 51/100[91/100].

Consider-the-opposite.

To make this decision, please first list two reasons why 51/100[91/100] might be too low[high] a rating for the *employee's performance this year*.

- Reason 1 _____
- Reason 2 _____

Now, indicate how would you assess your subordinate on a scale from 0-100:

APPENDIX D. SUB-GROUP ANALYSES FOR CHAPTER 3

Table 10*Sub-group analysis for the goal setting experiment*

	Managers			Employees		
	Replication	Low CTO	High CTO	Replication	Low CTO	High CTO
t-score	9.29***	3.39**	-.37	14.55***	4.17***	-3.39**
Cohen's <i>d</i>	1.42	0.50	.00	1.39	0.41	0.33
<i>n</i>	189	186	184	422	418	416
M(SD)						
Low	3.42 (2.13)	3.42 (2.13)		3.82 (3.30)	3.82 (3.30)	
High	18.80 (15.58)		18.80 (15.58)	25.80 (21.89)		25.80 (21.89)
Low CTO		4.69 (2.94)			5.41 (4.41)	
High CTO			17.97 (15.25)			19.48 (15.77)

*** $p < .001$; ** $p < .01$

Table 11
Sub-group analysis for the performance feedback experiment

	Managers			Employees		
	<i>Replication</i>	<i>Low CTO</i>	<i>High CTO</i>	<i>Replication</i>	<i>Low CTO</i>	<i>High CTO</i>
<i>t</i> -score	16.94***	2.58*	-6.95***	18.26***	1.63	-6.58***
Cohen's <i>d</i>	2.47	0.38	1.02	1.72	0.16	0.65
<i>n</i>	190	187	186	422	416	416
M(SD)						
Low	67.58 (9.46)	67.58 (9.46)		70.71 (10.38)	70.71 (10.38)	
High	88.36 (7.18)		88.36 (7.18)	88.20 (9.25)		88.20 (9.25)
Low CTO		70.77 (7.14)			72.13 (7.17)	
High CTO			79.61 (9.74)			82.49 (8.39)

*** $p < .001$; ** $p < .01$; * $p < .05$.

APPENDIX E: TOPIC LIST SEMI-STRUCTURED INTERVIEWS FOR CHAPTER 4

0. Introduction with information about the research and informed consent.

Information about the respondent:

1. Can you tell me something about yourself and your role within the municipality?

Example questions on views and experience

1. Which experience did you have with technology before the digital meetings?
2. What was your view on digitalisation in general?
3. What did you have to do to learn the practice of digitally meeting?

Example questions on changes

1. Which way did the change to digital meetings change the practice of council meetings? Why?
2. How did you experience the changes? Why?
3. What became more difficult because of digital meetings? Why?
4. What became easier because of digital meetings? Why?
5. What became more visible because of digital meetings? Why?
6. What became less visible because of digital meetings? Why?

Example question on environment

2. Does a difference exist between debating about simple and complex issues?

Example questions on differences between individuals

3. How do (other) council members experience these changes?
4. Some council members find meeting digitally very difficult, while others do not. Which factors do you think can explain this?

Example question on differences between municipalities

1. There are differences between municipalities, for example in how digital council meetings are experienced. What do you think can explain these differences?

Closing

1. Do you have anything to add?
2. I am looking for council members/municipality staff with diverse experiences with this subject. Do you know anyone I can contact?
3. Would you like me to keep in touch about the results?
4. Thank you

APPENDIX F: THE THEORETICAL DOMAINS FRAMEWORK

Table 12

The Theoretical Domains Framework domains and definitions (Cane et al., 2012; Michie et al., 2005)

Domain	Definition [possibly relevant constructs]
1. Knowledge	An awareness of the existence of something [procedural knowledge, knowledge of task environment].
2. Skills	An ability or proficiency, acquired through practice [skills, skills development, competence, ability, interpersonal skills, practice, skill assessment]
3. Social/Professional Role and Identity	A coherent set of behaviors and displayed personal qualities of an individual in a social or work setting [professional identity, professional roles, social identity, identity, group identity, leadership, professional boundaries, professional confidence, organisational commitment]
4. Beliefs about Capabilities	Acceptance of the truth, reality, or validity about an ability, talent, or facility that a person can put to constructive use [perceived competence, self-esteem, self-efficacy, self-confidence, empowerment, perceived behavioral control, professional confidence].
5. Optimism	The confidence that things will happen for the best or that desired goals will be attained [optimism, pessimism, unrealistic optimism, identity].
6. Beliefs about Consequences	Acceptance of the truth, reality, or validity about outcomes of a behavior in a given situation [outcome expectancies, anticipated regret, consequences].
7. Reinforcement	Increasing the probability of a response by arranging a dependent relationship, or contingency, between the response and a given stimulus [incentives, punishment, consequences, reinforcement].
8. Intentions	A conscious decision to perform a behavior or a resolve to act in a certain way [implementation intention, stages of change model]
9. Goals	Mental representations of outcomes or end states that an individual wants to achieve [goal setting, action planning, goal priorities]
10. Memory, Attention and Decision Processes	The ability to retain information, focus selectively on aspects of the environment and choose between two or more alternatives [tiredness, memory, attention, cognitive overload, attention, decision-making]
11. Environmental Context and Resources	Any circumstance of a person's situation or environment that discourages or encourages the development of skills and abilities, independence, social competence, and adaptive behavior [resources, environmental stressors, organisational culture, salient events, barriers and facilitators, person x environment interaction]
12. Social influences	Those interpersonal processes that can cause individuals to change their thoughts, feelings, or behaviors [social norms, group conformity, social comparisons, group norms, social support, power]
13. Emotion	A complex reaction pattern, involving experiential, behavioral, and physiological elements, by which the individual attempts to deal with a personally significant matter or event [fear, anxiety, affect, stress, depression, burn-out].
14. Behavioral Regulation	Anything aimed at managing or changing objectively observed or measured actions [self-monitoring, breaking habit, action planning]

APPENDIX G: EXPERIMENTAL CONDITIONS FOR CHAPTER 5

TOP OF FORM

G.1. EXPERIMENTAL CONDITIONS STUDY 1

We would like to ask you to judge the following scenarios. These scenarios are about managing back office employees in a municipality. Back office employees treat various requests from citizens, varying from providing documents to granting authorization. There is a trade-off between individual and collective interests per case.

Low in complexity practice

In a municipality a [computer, using an automated algorithm/a public manager, using his own judgment and an algorithm/a public manager using his own judgment] decides which travel reimbursement for commuting a back-office employee is entitled to.

How fair is the procedure by which it is decided by which travel reimbursement back office employees is entitled to?

High in complexity practice.

In a municipality a [computer, using an automated algorithm/a public manager, using his own judgment and an algorithm/a public manager using his own judgment] evaluates the performance of a back-office employee.

How fair is the procedure by which the performance of back office employees is evaluated?

G.2. EXPERIMENTAL CONDITIONS STUDY 2

The practices were presented randomly to participants.

Low in complexity practice.

The first practice is calculating pensions for Customer Service Officers. The calculation is based on a limited number of factors such as pensionable pay and inflation. The calculation is determined by law.

High in complexity practice.

The second practice is hiring Customer Service Officers. Hiring is based on a large amount of factors including fit with local governments values, fit within the team dynamics and employees' skills in service provision. The hiring process moreover consists out of multiple phases including the scanning of CVs and interviewing potential candidates. Opinions about which specific factors matter most differ.

Appendices

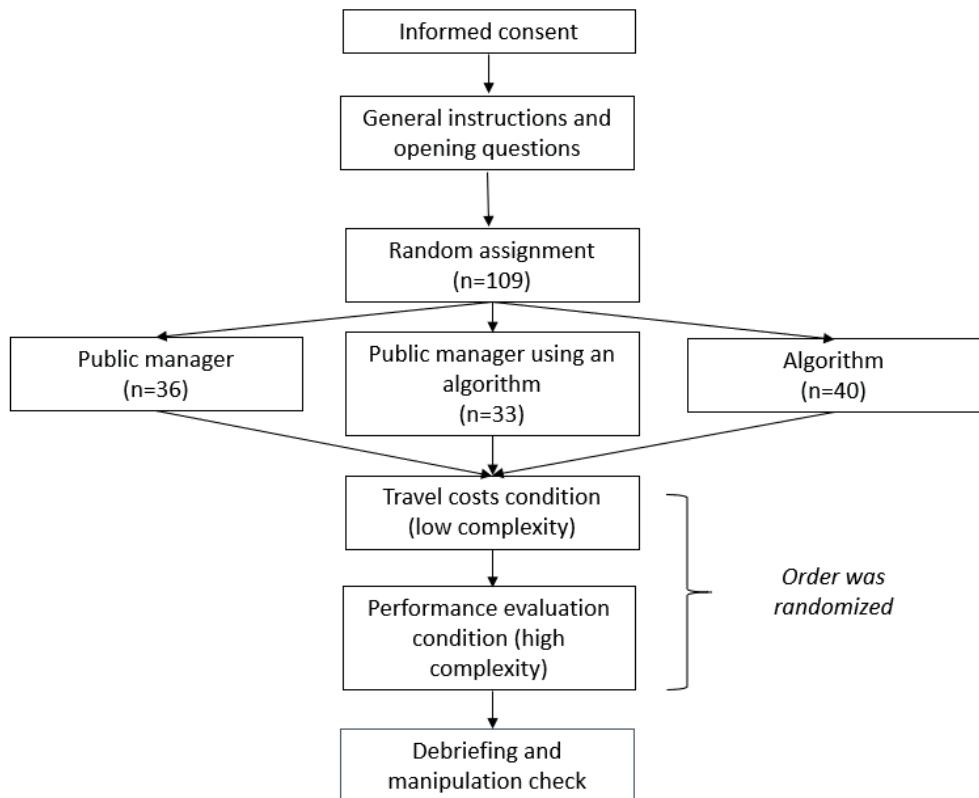
How fair is the procedure for [practice] for the Customer Service Officer when decisions are being made by:

- A public manager, using his own judgment.
- A public manager, using his own judgment and an algorithm.
- A computer, using an automated algorithm.

APPENDIX H: EXPERIMENTAL FLOW FOR STUDY 1 OF CHAPTER 5

Figure 9

Experimental flow for study 1



A

APPENDIX I. DEMOGRAPHIC DATA STUDY 1 AND STUDY 2 OF CHAPTER 5

Table 13*Descriptives and differences per group per condition for Study 1*

Variable	Public manager	Public manager using an algorithm	Algorithm	All
% female	57%	47%	55%	53%
% manager	33%	45%	40%	39%
Average year of birth	1977	1976	1977	1977
<i>Public sector</i>				
Healthcare	19%	24%	30%	25%
Education	22%	9%	18%	17%
Government	39%	27%	40%	38%
Other	19%	39%	13%	23%

Note. All differences between groups were tested through Chi-square tests, except the difference for age, which was calculated using an ANOVA. $p < 0.05 = *$

Table 14*Descriptives for Study 2*

Variable	All
% female	75%
% manager	43%
Average year of birth	43
<i>Public sector</i>	
Healthcare	16%
Education	41%
Government	20%
Other	20%

APPENDIX J – CO-AUTHOR STATEMENTS FOR CHAPTER 2 AND 3

The first author named is lead and corresponding author. We describe contributions to the paper using the taxonomy provided by Brand et al. (2015). *Rosanna Nagtegaal*: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original Draft *Lars Tummers*, *Mirko Noordegraaf* *Victor Bekkers*: Supervision, Writing - Review & Editing on all elements mentioned above.

A

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

This dissertation would not have been possible for the support and guidance of the people around me. First, I would like to thank my supervisors Lars, Mirko and Victor.

Lars, you have helped me since I was a student in the RESMA. I remember very well being in a meeting with you and Wouter Vandenabeele and not having a clue what statistical significance was or how to even spot it. At that time, you started explaining: 'When there is a star, that means it is significant'. You have taught me many things from scratch about science and the research process. Thank you so much for taking the time and effort to do that. After that, you hired me for a PhD and that led to this dissertation. During this PhD, you have been invaluable. Your enthusiasm and honesty have often inspired me to push a little further. Your trust has helped me to believe in myself and to get through the darker days, and our shared interests combined with your expertise have taught me so much of what I know.

Mirko, a.k.a. the Bob Dylan of Public Administration, as some have called you. At the beginning I remember being intimidated by you. I saw you as 'the big professor'. Now this has changed to 'the big friendly professor', as I have gotten to know you can be very supportive and kind. You have one foot in the academic world and one foot in practice and it did not matter what the subject was, you always knew something about it, or knew someone who knew something about it. You have also taught me many things about how to communicate my work and your vision helped me tremendously in my progress.

Victor, you have given me guidance in a very modern way. Far before the COVID-19 crisis we have interacted in a hybrid form. I always tell everyone how smart you are. Your comments are often about the weaknesses of my work of which I hoped they would go undetected. But no, when you are in the meeting, I know they are going to be mentioned. This has helped me a lot. On top of that, your skills to relate articles to the 'debate' in public administration and the eternal call for the research question have kept me sharp.

Lars, Mirko and Victor, thank you so much. You will always have my respect and gratitude.

This research was also supported professionally by the many people who helped me like the people from the PSI-CO project, including Koen Verhoest and Tom Langbroek, the people at the BIN NL network, especially Femke Breukels and Manoe Mesters and the people at Microsoft including Florine Segers, Sabine Hess and Tony Crabbe. I would also like to thank the council members and politicians who shared their experiences with me and of course the anonymous respondents that provided the data for the survey experiments. This research also benefitted greatly from the feedback and experiences I obtained at conferences such as EGPA, IRSPM and PMRC.

I also want to thank my other colleagues at USBO and elsewhere, including Rianne Dekker, Scott Douglas, Albert Meijer, Barbara Vis, Daniëlle van Loo, Stephan Grimmelikhuijsen,

Wouter Vandenabeele, Sjors Overman, Robin Bouwman, Amandine Lerusse and many more. Everyone I met is so smart and driven to make it work. It is really astonishing. It was good to be in your presence. This is true as well for my fellow PhDs including Alex, Arend, Erik-Jan, Rutger, Marlot, Igbal, Arvie, Evelien, Maurits, Marij, Lauren, Lars, Joram, Jo, Noor and Isa, as well as all the new PhDs. It has been a pleasure to share the PhD experience with you and to learn from you on a personal as well as professional level.

Then my sweet paranympths Julia and Marija. Julia, during our trip to Lisbon I soon spotted how cool you were and that you had friend potential. This turned out to be true. You have such a big heart and are always seeking to be a good person. You have been there in many fun, but also in some darker moments. I will never forget that. You bring style to whatever you do, whether you are presenting a lecture or you are shining on the dancefloor. You are also the most all-round scholar I know. I was already impressed by you in the RESMA. You are just good at everything! You know so much about statistics, about connecting practice to theory, about presenting, about teaching and the list goes on. I've always looked up to you as a person and am so proud to call you my paranympth.

Marija, when we met, you've just beaten me a job interview trajectory. I must admit, I was a little bit jealous at that time. But that quickly turned into happiness as I saw how great you are. It was just impossible for me not to like you. The way you tackled the topic of accountability, your knowledge about obscure topics, your dry humor, cooking skills and work ethic never cease to surprise me. You are also one of the sweetest and most hospitable persons out there. I feel very lucky and grateful to call you my paranympth and look forward to many more jokes, dinners and boardgames with you!

I also want to thank my other friends. Simon, your humor, kindness, knowledge about pretty much anything and the ability to persevere, either on the boulderwall or in life, are great. To many more bouldersessions and dinners!

Idwer, from our dancing in Lisbon, to our dancing basically everywhere where we have the chance, to our weekly boulder-and-biertje-sessions, to philosophical conversations in art galleries. It is a true gift to be your 'mattie'. Never lose your curiosity and openness to life!

Edu, Julia, Andres, Nathalie and Ivan, la familia, you are all amazing and I am so grateful that I have gotten to know you. Thanks for all the laughter, support and way too many shots. I bet there are much more to come. Of course also a special mention to Hugo, Dan and Geesje.

Eva, you teach me new things every time I see you about life and how to approach it. Already for more than 15 years, you are the Ying to my Yang. I cannot wait to enjoy many more years being your friend!

Yvette, Michiel en Erik, my academic career started in the best of the best places, on the Pieter Breughelplein. Every time I think of that time it brings a smile to my face. Breughels for life.

Acknowledgements

And of course I cannot forget my sweet family. Dear Jasper, I have often thought about you when writing this dissertation, and I hope this contributes to better decisions by for instance doctors. Dear Nina, Jet, Ingrid and Peer, liefde voor altijd! Lieve oma Nel en oma Miek, ik vind het heel leuk dat jullie in mijn leven zijn en geniet van onze bezoekjes. Ik hoop op nog veel meer bezoekjes en leuke gesprekken in de toekomst!

Mom and dad, there is no way that I can thank you enough for what you have done for me. I have been very lucky with you as parents and I am very grateful we have such a good relationship. You always support me, no matter what I do. You have taught me a lot about life and also about the topic of my thesis, by sharing your experiences in your own organizations. Thank you for everything.

Wilma, I am very happy to have you as a bonus-mom. You are so sweet and down to earth. I am learning a lot from you.

Evy, we are carved from the same wood. We are always connected which I love about us. I value your perspective on life and how you can help me focus on what really matters. Kosta, je inspireert me elke keer als ik je zie door jouw open blik naar de wereld. Seb, my wish came true to have you as a part of our family.

Laura and Jesús, gracias por vuestra hospitalidad, por aceptarme en vuestra familia y mostrarme tantas cosas buenas de su cultura. Estoy deseando muchos más buenos momentos juntos en vuestra hermosa España.

Gabi, you are the purest human being I have ever met and have a true heart of gold. Makes sense, seeing that you are Spanish. Jokes aside, I love you and all that you have done for me. You will forever have a place in my heart.

A

ABOUT THE AUTHOR

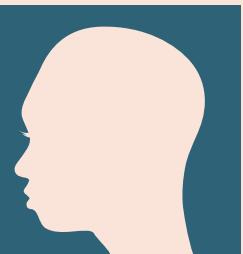
ABOUT THE AUTHOR

Rosanna Nagtegaal was born in 1990 in Vlissingen. She graduated from her high school as the top of her class, while being active in her school's participation council. Afterwards, she went on to study Public Administration at Tilburg University. During her bachelor, she became an UniPartner at UniPartners Tilburg. In 2014 she started the two-year research master program in Public Administration and Organizational Science. This program is offered by Utrecht University, Erasmus University Rotterdam, Tilburg University and the Vrije Universiteit Amsterdam. During her master, she became a student assistant for the 'Learning from Innovation in Public Sector Environments' (LIPSE) project. In 2016 she graduated cum laude. The research of her master thesis became national news by making the frontpage of the newspaper Trouw and led to parliamentary questions.

Afterwards, she began her PhD at the Utrecht School of Governance (USG). During her PhD she was a chair for the AIO-platform, an ambassador for the BIN NL network, a teacher and a supervisor. She also collaborated on the 'Public Sector Innovation through Collaboration'(PSI-CO) project. She published her work in journals such as Bestuurskunde, the Journal of Behavioural Public Administration, Public Administration Review and Government Information Quarterly and presented her work at conferences such as EGPA, IRSPM, PMRC, WINK and NIG. She is also a chair of the Netherlands Institute of Governance (NIG) 'Behavioural Public Administration' panel.

Currently Rosanna is working as a post-doc on the 'Together to Work' (Dutch: 'Samen aan het werk') project. She researches how behavioural sciences can be used to influence employers' hiring decisions about people with disabilities.

A



The rise of behavioral public administration demonstrated that we can understand and change decision-making by using insights about heuristics. Heuristics are mental shortcuts that reduce complex tasks to simpler ones. Whereas earlier studies mainly focused on interventions such as nudges, scholars are now broadening their scope to include debiasing, and psychological theories beyond heuristics. Scholars are moreover shifting their attention away from citizen-focused interventions to public sector worker-oriented interventions, i.e. the very people who are expected to nudge society. This dissertation seeks to explore how behavioral sciences can facilitate understanding and support decision-making across the public sector. We present four studies that investigate a range of behavioral theories, practices, issues and public sector workers. This dissertation shows that when handling heuristics in the public sector, we need to take into account the institutional and situational settings, as well as differences between public sector workers. The results of this dissertation can be used by practitioners and academics to understand and support decision-making in public sector contexts.