

Marija Aleksovska

UNDER WATCHFUL EYES

Experimental studies
on accountability and
decision-making
behavior in the
public sector



Under Watchful Eyes

Experimental Studies on Accountability and Decision- Making Behavior in the Public Sector

Marija Aleksovska

Doctoral committee:

Prof. dr. Mark Bovens

Dr. Madalina Busuioc

Prof. dr. Elianne van Steenbergen

Prof. dr. Wouter van Dooren

Prof. dr. Gregg G. Van Ryzin

Prof. dr. Barbara Vis

Under Watchful Eyes Experimental Studies on Accountability and Decision-Making Behavior
in the Public Sector

ISBN: 978-90-9034945-9

Cover design: Filip Velkovski

Layout and design: ProefschriftMaken

Printing: ProefschriftMaken | www.proefschriftmaken.nl

Copyright © 2021 Marija Aleksovska

All rights reserved. No part of this publication may be reproduced, stored in a retrieved system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing from the proprietor(s).

Under Watchful Eyes

Experimental Studies on Accountability and Decision-Making
Behavior in the Public Sector

Bij Waakzame Ogen

**Experimentele Studies over Verantwoording en Besluitvormingsgedrag
in de Publieke Sector**

(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 17 september 2021 des middags te 4.15 uur

door

Marija Aleksovska

geboren op 9 augustus 1989
te Skopje, Macedonië

Promotor:

Prof. dr. T. Schillemans

Copromotor:

Dr. S.G. Grimmelikhuijsen

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Acknowledgements

Writing a PhD thesis has been a process of learning for me, not only about public accountability and experimental methods, but also who I am, and who I want to be as a researcher. This process was greatly shaped by the people around me, who have offered me their knowledge, guidance, support, and brightened my days with their presence.

First and foremost, I would like to thank my supervisors Prof. dr. Thomas Schillemans and dr. Stephan Grimmelikhuijsen for guiding me through the academic world and the PhD process. They were more than generous with their time and knowledge, and their feedback on my work was invaluable. But their guidance extended much further than feedback on my work, as they helped me navigate the academic world of conferences and journals, and showed me how to find my place in it. Thomas, I deeply appreciate all the effort and care you have put towards my academic development. Stephan, your enthusiasm and useful advice made this PhD process smoother than I ever imagined. I am forever grateful for everything you have done for me. Thank you.

The research conducted for this dissertation would have not been possible without the funding of the Dutch Research Council (NWO), as well as the help of the kind professionals at Bureau Inspectieraad, and the Dutch School for Public Administration (NSOB), and notably Kaspar van den Ham, Rob Kerstens, Prof. dr. Ian Leistikow, Hadewych van Kempen, Ytzen Renema, Myrte Ferwerda, Jacob Smit, Erik Voerman, Rémi Langenberg, Mieke den Bieman, Rob Velders, Antonia Sattlegger, and Kees van der Wel.

The colleagues at Utrecht School of Governance warmly welcomed me when I first joined the department in 2016, and made me feel like I was home ever since. I greatly enjoyed being a part

of the Calibrating Public Accountability team, where I had the chance to work with and learn from the brilliant Sjors, Lars and Meike.

My fellow PhDs have made my dissertation writing days much more bright and colorful. I am very grateful to have had their advice in the last few years, but also to have had them as colleagues and friends. Joram, Jo, Lauren, Alex, Marij, Erik-Jan, Arend, Maurits, Olga, Minou, Marlot, Igbal, Rutger, Evelien, Isa, and the new generation of PhDs, thank you for being part of my PhD journey!

Rosanna and Julia, your friendship has been perhaps the most valuable outcome of this PhD process. I feel incredibly fortunate to have found you both at USG, and to have formed our Kelly, Michelle, and Beyoncé bond with you.

My family has selflessly supported my academic ideas and ambitions, even though that meant being away from home and away from them for almost a decade. I am eternally grateful for that.

Мама, тато и Дарко, фала ви за целата поддршка и помош сите овие години!

Without the moral and emotional support of my friends, these PhD years would have been much more difficult and dull. Ana, Katina, Elena, Andrijana, Maja, Vlada, Pupsi, Alex, Milka, Joey, Ross and Kalin, thank you! I am especially grateful to my friend Filip Velkovski for doing an amazing job in turning my vague ideas about the cover of this dissertation into a very cool reality.

Inga I am very glad to have met you in Leiden and fortunate that both of our paths led to a PhD in Utrecht. Hoping for many more years of brainstorming ideas together, sharing academic memes, and fun beach outings!

Lastly, I am thankful to have had Simon by my side all these years. For his love and support, I am forever grateful.

Table of contents

List of tables.....	v
List of figures.....	v
1. Introduction: A behavioral perspective on public accountability.....	1
1.1. Public accountability.....	4
1.2. Accountability mechanisms in the public sector.....	6
1.3. The research gap: A lack of systematic and causal identification of the behavioral effects of accountability mechanisms.....	12
1.4. Public accountability through a behavioral lens.....	14
1.5. Research questions.....	16
1.6. Scientific relevance: Integrating behavioral theories and methods.....	18
1.7. Societal relevance: Optimizing accountability.....	21
1.8. Outline of dissertation.....	23
2. Accountability in the behavioral sciences: Lessons for public administration.....	25
2.1. Introduction.....	26
2.2. Accountability mechanisms in the public administration literature.....	29
2.3. Systematic review: Methods and data.....	34
2.4. Effects of accountability mechanisms in behavioral research.....	38
2.5. Research agenda for accountability in public administration.....	50
3. Accountable for what? The effects of specifying the accountability standard.....	57
3.1. Introduction.....	58
3.2. Theoretical basis for understanding behavior under accountability pressure.....	61
3.3. The social contingency model of accountability.....	64
3.4. The case for general accountability standards.....	67
3.5. The case for specific accountability standards.....	68
3.6. Methodological approach.....	70
3.7. Study 1.....	76
3.8. Study 2.....	83
3.9. Discussion.....	88
3.10. Conclusion.....	93
4. Accountable to whom? Understanding what makes the accountability forum matter.....	95
4.1. Introduction.....	96
4.2. The unavoidable multiple accountabilities: Overview of the literature.....	99
4.3. Competing views on accountability.....	101
4.4. Study design and data.....	106
4.5. Results.....	112
4.6. Discussion.....	117
4.7. Conclusion.....	122

5. Managing multiple accountabilities: The effects of facing multiple forums and conflicting demands	123
5.1. Introduction	124
5.2. The multiple accountabilities problem in the literature	127
5.3. Disentangling the elements of the concept of multiple accountabilities and their effects	129
5.4. Decision-making strategies to cope with multiple accountabilities pressures	132
5.5. The accountability of regulators.....	138
5.6. Experimental design and participants	139
5.7. Analysis and results.....	144
5.8. Discussion	148
5.9. Conclusion.....	153
6. Conclusion: Understanding accountability through a behavioral lens	155
6.1. Why we need to understand the behavioral consequences of accountability in the public sector	156
6.2. Accountability and decision-making behavior: Universal and context-dependent effects	160
6.3. Limitations of the presented research.....	163
6.4. Behavioral science and public sector accountability: A mixed blessing	164
6.5. Towards a behavioral theory of public accountability?	172
References.....	177
Appendices.....	197
Appendix Chapter 2	198
Appendix Chapter 3	235
Appendix Chapter 4	244
Appendix Chapter 5	260
Dutch summary.....	273
About the author	279

List of tables

1.1 Overview of the empirical chapters of the dissertation	24
2.1 Studied characteristics of accountability mechanisms.....	47
3.1 Overview of measures.....	75
3.2 Manipulation checks results.....	78
3.3 Effects of accountability standard specification on decision-making effort from Study 1.....	80
3.4 Effects of accountability standard specification on decision impartiality from Study 1	82
3.5 Effects of accountability standard specification on decision-making effort from Study 2.....	86
3.6 Effects of accountability standard specification on decision impartiality from Study 2	87
4.1 Overview of the attributes and their levels	108
4.2 Overview of subject characteristics	110
5.1 High-effort strategies results from two-way ANOVA and chi-squared analysis	147
5.2 Low-effort strategies results from two-way ANOVA analysis	148

List of figures

1.1 Elements of an accountability relationship	7
2.1 Frequency of conducted experiments over time	37
2.2 Type of experimental task frequency.....	38
2.3 Effects of accountability on decision-making.....	41
2.4 Effects of accountability on behavior	43
2.5 Effects of accountability on decision-making and behavior outcomes	46
3.1 Overview of the experimental design	72
4.1 Accountability forum prioritization – pooled sample.....	114
4.2 Accountability forum prioritization per country - matched sample.....	117
5.1 Flowchart of the experiment	141
6.1 The research focus of the dissertation.....	159

1. Introduction: A behavioral perspective on public accountability

The 1986 Challenger tragedy has become a classic case of public accountability gone wrong (Romzek & Dubnick, 1987). On January 28, 1986, the space shuttle Challenger exploded only 73 seconds after liftoff, killing all of its crew members on board. The cost of the disaster was not only in human lives, but also in the high price paid to replace the shuttle, delays in the deployment of commercial and military satellites, as well as damage to the National Aeronautics and Space Administration's (NASA) professional reputation (Heimann, 1993). The Rogers Commission, tasked with investigating the accident, concluded that there were two key contributing causes to the disaster: a technical one; and a managerial one. The technical cause referred to the failure of the O-ring seals, which were not designed to handle the unusually low temperatures at the time of the launch. The managerial cause referred to problems in communication that resulted in the failure of the management to stop the launch, even when alarms about the performance of the O-ring seals under low temperatures were raised by engineers. Public administration scholars Romzek and Dubnick (1987), however, have protested that these two causes do not provide a sufficient account of the factors that led to the disaster, and argued for the inclusion of an analysis of the institutional context in which NASA operated. In their seminal study for accountability research, they contend that the Challenger disaster was in part a result of NASA's efforts to manage the multiple diverse expectations of its various accountability-holders in its institutional environment (Romzek & Dubnick, 1987).

As a public sector agency, Romzek and Dubnick argue, NASA was faced with legitimate expectations stemming from its political, legal, bureaucratic as well as professional accountability (1987). In the period leading up to the disaster, the agency no longer enjoyed the great political and public support, as well as financial comfort, of its earlier days. It was, therefore, under great pressure to perform in order to justify its large operational costs and the backing received by its

political principals and other actors. As a result, NASA's top managers set up an ambitious launch schedule, which was proving difficult to meet. The political pressure that NASA's top management was experiencing was transmitted down to lower levels of the bureaucratic hierarchy. The middle managers thus felt the burden not only of the expectations of their bureaucratic superiors, but also of the expectations of the agency's political principals. It was in this context that NASA's middle management decided that it was safe to proceed with the planned launch of the shuttle, and not to delay the process by resolving the technical issues the engineers raised. The middle management thus prioritized addressing the expectations stemming from the agency's political and bureaucratic accountability, at the expense of its professional accountability, ultimately leading to the disaster.

This influential analysis of the Challenger accident encouraged numerous investigations of the accountability forces operating in the public sector in the decades to come. The simultaneous operation of multiple accountabilities, in particular, has garnered a lot of interest among public administration scholars (for example Koppell, 2005; Romzek & Ingraham, 2000; Schillemans & Bovens, 2011). But while the various accountability mechanisms have been carefully traced and categorized, their consequences on the decision-making behavior of individual public sector actors have been subjected to far less scholarly scrutiny (Schillemans, 2016; Yang, 2012). We thus have far less insight into why NASA's middle management decided to give priority to the demands of their political principals and hierarchical superiors, as opposed to the ones of the technical experts. The decision-making behavior of public sector actors under the pressure of accountability is therefore a central subject of investigation in this dissertation.

1.1. Public accountability

Accountability has been said to be a “golden concept” that no one can be against (Bovens, 2007, p. 448). It is considered as a cornerstone of modern democratic regimes of governance, and a central mechanism for steering behavior towards meeting standards of appropriateness and achieving performance goals (Schillemans, 2015). The operation of accountability mechanisms in the public sector has thus often been justified with great promises for desirable outcomes (Dubnick, 2005).

The last several decades of accountability scholarship, however, have shown that the reality of their effects appears to be much more complicated. A wide range of consequences have been attributed to the workings of accountability mechanisms, many of which signal problems and unintended consequences. Too many accountability requirements have been found to be just as problematic as too few (Halachmi, 2014; Mulgan, 2014). Holding public sector actors accountable for the achievement of particular performance targets has inspired more “creative” ways of achieving those targets, instead of actual improvements in the quality of work (Bevan & Hood, 2006; Lægreid & Neby, 2016). Having political actors, citizens, professional groups and other stakeholders simultaneously act as account-holders has been argued to improve the ability to hold power to account (Schillemans, 2010; Willems & Van Dooren, 2012), but it can also confront public sector actors with irreconcilable expectations (Romzek & Dubnick, 1987; Thomann et al., 2018). It therefore appears that accountability mechanisms do not always simply lead to better governance outcomes. Their effects are many, and vary depending on the characteristics of the accountability mechanisms and the context in which they are enacted. To be able to instrumentally use them to achieve goals of better performance and professional conduct, we must first understand the way they operate, and the results they produce.

While public sector accountability mechanisms primarily operate at the organizational level, their effects reverberate through the various segments and individuals within organizations, as the case of the Challenger tragedy illustrates. Through board meetings, discussions in the media, or contacts with clients, the pressures of accountability expectations are felt by *individual* public sector managers and employees alike.

Felt accountability, as psychologists have dubbed this (Hall et al., 2017), affects one's reasoning about the problem at hand, the calculations of priorities, the willingness to invest additional effort, or to take risks, and subsequently the decisions one makes and the behaviors one performs in the workplace (Lerner & Tetlock, 1999). These individual-level decisions and behaviors in turn shape the organizational outcomes. The meso and macro-level accountability pressures, external to the organization, are thus channeled through the micro-level within the organization, before they return meso or macro-level outcomes (Overman et al., 2020). To understand meso and macro-level outcomes in the public sector, it is therefore crucial to learn how the individual-level decision-making behavior is shaped by the various accountability mechanisms.

This dissertation aims to take a step in the direction of a closer understanding of the effects of accountability mechanisms on individual decision-making behavior in the public sector. Decision-making behavior aims to capture the choices individuals make when presented with multiple alternative courses of action, but also the process of doing so. It, thus, refers not only to the decision made, but also to the decision-making process in which the decision was reached. To investigate the effects of accountability on decision-making behavior, this dissertation draws on the rich experimental and theoretical work on accountability developed in the behavioral sciences (Hall et al., 2017; Lerner & Tetlock 1999), and applies it to the public sector context.

This introductory chapter lays out the rationale behind the research conducted in the dissertation. First, it examines the knowledge regarding the operation of accountability mechanisms in the public administration literature and identifies the research gap that this dissertation targets. Second, the behavioral approach to accountability taken to address the stated gap is then explained. Third, the sub-questions guiding the examination of the central research question are outlined and explained. Fourth, the scholarly and practical relevance of the dissertation is presented, and finally, an outline of the following chapters is provided.

1.2. Accountability mechanisms in the public sector

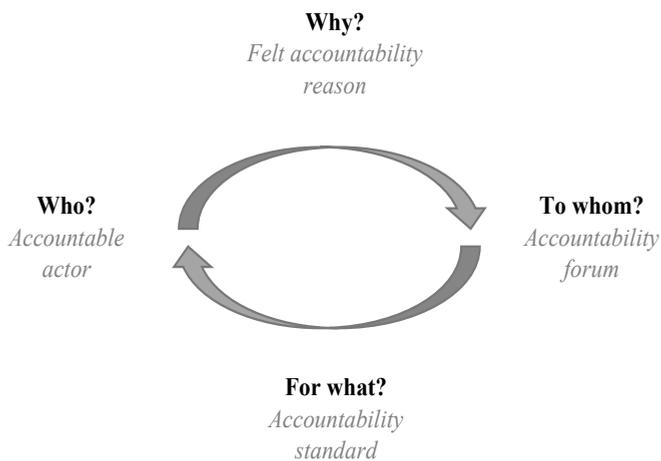
Before introducing the behavioral approach taken in this dissertation, it is necessary to look at the way accountability mechanisms are currently identified and discussed in the public administration literature. To do so, I follow Bovens' classic definition of accountability as: "*a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences*" (2007, p.447, 2010, p.951). Accountability is, thus, a mechanism capturing the institutionalized relation between two actors¹.

To describe a particular accountability mechanism, one must answer four questions: *who, to whom, why* and *for what*. The *who* question identifies the actor who is obliged to give account, while the *to whom* one captures the identity of the account-holder, also known as an accountability forum. The *why* question explains why the actor is compelled to give account for its actions, and the *what*

¹ This definition leaves out the normative understanding of the concept that sees accountability as a standard for evaluation of behavior, which is sometimes adopted in the literature (Bovens, 2007, 2010).

question concerns the accountability standard, or what the actor is being held accountable for (Bovens, 2007, pp. 454-455). The elements of an accountability relationship are displayed in Figure 1.1. In describing the accountability mechanisms found in the public sector, these four questions receive many different answers, and form a sizeable number of combinations. Thus, accountability mechanisms are hardly a uniform phenomenon, taking many different shapes (for some classifications see Behn, 2001; Bovens, 2007; Romzek & Dubnick, 1987).

Figure 1.1 *Elements of an accountability relationship*



Who? Both individuals (such as civil servants), and collectives (such as organizations), take the role of actors in accountability relationships (Bovens, 2007). On the individual level, scholars have investigated the accountability mechanisms in which street-level bureaucrats (e.g. Hwang, 2016; Murphy & Skillen, 2015; Thomann et al., 2018), public professionals (e.g. Ewert, 2020; Metzl & Speckbacher, 2015), and also public managers (e.g. Schillemans, 2015; Schwabenland & Hirst,

2020) adopt the roles of accountable actors. The majority of the analyses of accountability mechanisms in public administration, however, focus on the organizational level (e.g. Bach et al., 2017; Brodtkin, 2008; Koppell, 2005; Overman et al., 2015; Schillemans, 2010, 2011; Verschuere et al. 2006).

Of key interest in these analyses are the characteristics of the accountable actors, regardless of whether this pertains to individuals or collectives. These include the tasks the actors handle in their professional capacity, the domain in which they work, the autonomy they enjoy in their work, and the institutional and legal arrangements which allow the actors to perform their public tasks (e.g. Bach et al. 2017; Verschuere et al., 2006; Thomann et al., 2018; Schillemans et al., 2020). These characteristics are often coupled with observed or reported patterns of account-giving, thereby mapping the accountability arrangements in particular domains of the public sector. In some instances, scholars have assessed multiple organizations, domains and even countries, and thus focused on outlining patterns of similarities and differences in account-giving, by means of comparison (e.g. Bach et al., 2017; Jantz et al., 2018; Verschuere et al., 2006). In other studies, scholars have focused on particular cases, capturing a single event, organization, or group of civil servants, and focused on teasing out the interaction between the accountable actors and the accountability pressures they are subjected to, in terms of behaviors (e.g. Hwang & Han, 2017; Romzek et al., 2012; Schillemans, 2015), and observed outcomes (e.g. Kim, 2018; Mills & Koliba, 2015; Romzek & Dubnick, 1987).

Why? There are a number of reasons why a public sector actor would feel compelled to give an account to an accountability forum, and they are closely related to the nature of the relationship

between the two (Bovens, 2010). One reason to give an account would be a formal obligation to do so, which primarily characterizes hierarchical and principal-agent relationships (Jarvis, 2014). A formal obligation to give account could also arise as a result of legal rules or professional standards (Schillemans, 2008; Scott, 2000). Public sector actors, however, often feel an obligation to give account to particular forums even though they are not formally required to do so (Romzek et al., 2012). This type of voluntary account-giving is commonly directed towards societal actors, the media, and the general public, and considered to be motivated by norms of appropriateness, but also strategic considerations (Karsten, 2015; Koop, 2014).

The drivers of account-giving have been explored in more detail in the theoretical work on public accountability. Thus, scholars have discussed the root motivations for account-giving, both when this is formally required from the public sector actor, and when it is not. Since accountability forums could potentially apply formal and informal sanctions to the actor when their expectations are not met, avoiding punishment is considered to be a strong motivating factor for account-giving. In the absence of formal requirements, account-giving is considered to be motivated by moral regards, as well as shared norms and values (Bovens, 2007; Karsten, 2015; Koop, 2014). Recent theoretical accounts have highlighted that actors also engage in account-giving in order to actively shape their reputation, and through that their space for strategic action. Account-giving is thus considered not only as a defensive tool, but also as a tool for empowerment of the accountable actor (Busuioc & Lodge, 2016, 2017).

For what? Public sector actors can be held accountable for a range of aspects of their work. These include financial workings, adherence to legal regulations, following established procedures of

work, attaining professional standards, achieving performance goals, responsiveness to political and public expectations, as well as maintaining appropriate behavior (Bovens et al., 2014; Dubnick, 2005; Romzek & Dubnick, 1987). Commonly, public sector actors are held accountable for several of these aspects of their work simultaneously.

How the content of the accountability standard is defined, and what consequences it produces, has been a subject of rich debate among scholars and practitioners. Specifically, the effects of holding one accountable for following particular procedures, as opposed to for delivering specific outcomes, has been discussed extensively, often in the context of New Public Management reform (Ossege, 2012; Romzek, 2000). The setting of specific performance targets has been a highly problematized element of this reform, as it has been argued to stimulate gaming behavior, instead of improved performance (Bevan & Hood, 2006; Hood, 2006; Lægreid & Neby, 2016). The accountability standard has also been discussed in terms of sheer amount of work it requires to be satisfied (Schillemans, 2016). Scholars have argued that here accountability in fact negatively affects performance as it takes away resources that public sector actors have to invest into fulfilling core tasks (Behn, 2001; Dubnick, 2005). In addition, the expectations set by accountability mechanisms limit the agency of actors to foster responsible judgment, and often introduce perverse incentives to work towards better self-presentation, as opposed to better performance, leading into a so-called “accountability paradox” (Behn, 2001; Jos & Tompkins, 2004).

To whom? Accountable actors often face the “problem of many eyes” (Bovens, 2007, p.455), which means that their work is (simultaneously) scrutinized by a range of accountability forums. These include hierarchical actors, such as parent departments and ministries, oversight bodies,

such as boards, auditors and inspectorates, but also professional associations, the media, societal stakeholders, clients, and the public at large (Ewert, 2020; Hupe & Hill, 2007; Willems & Van Dooren, 2012). The precise constellation of forums the accountable actor faces, and perhaps their relative importance for the actor, will depend on the positioning of the accountable actor in the public domain, as well as the nature of its activities.

This multitude of forums that the accountable actor faces, often at the same time, has been a subject of great scholarly interest. Some theoretical work has deemed this phenomenon primarily positive, since there are reasons to believe that it contributes to the quality of the governance regime (Schillemans, 2010; Willems & Van Dooren, 2012). The empirical investigations of the phenomenon, however, display much more varied findings. A sizeable part of the empirical literature portrays the simultaneous account-holding of multiple forums as problematic, as it introduces dilemmas and difficulties in the work of the public sector actors (Thomann et al., 2018; Ewert, 2020). This is essentially due to the diverging logics the various accountability forums follow in evaluating the actor's work, and the conflicts that sometimes emerge between the expectations they put forward to the actor (Kim, 2005; Klingner et al., 2002). In some cases, the consequences are found to be catastrophic, due to the inability of the public sector actors to assign priorities to the demands of the different accountability forums (Koppell, 2005), or due to the assignment of misguided priorities, like in the case of the Challenger disaster (Romzek & Dubnick, 1987). Not all empirical investigations find multiple accountabilities to be very problematic, however. In many instances, public sector actors have been found to develop strategic coping mechanisms to respond to the demands and expectations of their accountability forums, and have successfully integrated those into their workflow (Hwang, 2016; Hwang & Han, 2017; Schillemans, 2015; Schillemans & Bovens, 2011).

1.3. The research gap: A lack of systematic and causal identification of the behavioral effects of accountability mechanisms

The last few decades of accountability research have provided an excellent overview of the accountability landscape in which public sector actors operate. We thus have a good insight into the different types of accountability mechanisms in the public sector, as well as useful tools to categorize (Behn, 2001; Bovens, 2007; Koppell, 2005; Romzek & Dubnick, 1987) and measure them (Bovens et al., 2008; Brandsma & Schillemans, 2013; Han & Perry, 2020b; Overman et al., 2020; Rock, 2020). In addition, we have numerous analyses of their functioning in particular domains (e.g. Mills et al., 2018; Romzek & Johnston, 2005; Thomann et al., 2018) and in key events (e.g. Gregory, 1998; Romzek & Dubnick, 1987; Romzek & Ingraham, 2000), as well as accounts of experiences and strategies for coping with the accountability pressures of various individuals within the public sector (e.g. Ewert, 2020; Hwang & Han, 2017; Schillemans, 2015). These are indispensable for understanding the way accountability processes unfold in the realities of the public sector.

The extant accountability literature suggests a number of links between the functioning of particular accountability mechanisms and the behaviors of public sector actors. These are, however, primarily deduced from past events and individual accounts, and rarely directly observed or tested (Dubnick, 2005; Yang, 2012, 2014). For example, in the analysis of the Challenger tragedy (Romzek & Dubnick, 1987), the effect of the accountability pressures on the behavior of NASA's management was deduced from accounts on the event itself and the inner-workings of the agency. Some suggested links between accountability and behavioral outcomes have not been empirically tested at all, and yet they appear to carry a lot of weight in accountability scholarship. For example, the conflict between accountability and performance which constitutes the often-

discussed “accountability paradox” (Behn, 2001; Jos & Tompkins, 2004) has been only elaborated theoretically, even though it captures concrete behavioral assumptions and measurable outcomes (Dubnick, 2005; Yang, 2014). Rigorous investigations of the causal effect between accountability structures and individual agency thus have not yet been developed. It is thus unclear what are the powers and possibilities of accountability mechanisms to shape the behavior of individuals and subsequently lead to particular meso or macro-level outcomes (Yang, 2012, 2014). In order to understand and anticipate the effects of particular accountability arrangements, drawing clear causal links is of key importance.

The lack of systematic and causal investigation of the effects of the various accountability mechanisms is the research gap this dissertation addresses. The focus is placed on two of the four elements of the accountability mechanism, namely the accountability standard and the accountability forum. The accountability standard is a particularly relevant element for further investigation, due to its high practical significance. It is namely the most pliable of the four, and thus most easily subjected to reform. This is also partly the reason why it appears so frequently in scholarly and policy discussions. The accountability forum, in contrast, is highly inflexible. Public sector actors have accountability obligations to a wide range of forums, and it is part of their work to continuously manage and respond to their expectations (Busuioc & Lodge, 2017). Understanding the way they do so both successfully and unsuccessfully is crucial for making sense of public sector outcomes. As accountability mechanisms rarely operate in isolation, their simultaneous effects are also investigated. This is particularly relevant due to the mixed views and evidence regarding their effects.

1.4. Public accountability through a behavioral lens

Over the last decade or so, scholars within the discipline of public administration have displayed renewed interest in the micro-level, and thus focused on investigating the workings of individual attitudes and behaviors of civil servants and citizens (Grimmelikhuijsen et al., 2017). They thus have integrated concepts, theoretical models and methodological approaches from psychology and the other behavioral sciences into the study of public administration (Tummers, 2020). This behavioral perspective has gained a lot of ground in recent years that is now considered as a distinct approach within the discipline of public administration, referred to as *behavioral public administration* (Grimmelikhuijsen et al., 2017; Tummers, 2020).

The interdisciplinary cross-fertilization between public administration and the behavioral sciences that the behavioral public administration approach promotes has proved useful for advancing the understanding of a number of problems in the public sector. These include transparency (Grimmelikhuijsen & Meijer, 2014), administrative burden (Christensen et al., 2020), performance information (James et al., 2020), and many more. This behavioral perspective has started to emerge also in the study of public accountability. Scholars have integrated behavioral elements in their theoretical and conceptual models (Han & Perry, 2020a; Schillemans, 2016), as well as in their empirical designs (Bouwman et al., 2018; Ossege, 2012; Schillemans et al., 2020), aimed to capture the functioning of accountability mechanisms in the public sector. The “behavioral turn” of the study of public accountability, though, is still only at its nascence.

The behavioral sciences, and in particular social psychology, have a rich tradition of investigating the effects of accountability (Hall et al., 2017; Lerner & Tetlock, 1999). Within these disciplines, scholars have developed theoretical models and engaged in extensive experimental work to test

their propositions. The key theoretical model in this research tradition is the *social contingency model of judgement and choice* which was developed by Philip Tetlock (1992, 1999, 2002).

The social contingency model of judgment and choice regards individuals as “intuitive politicians”, who adapt their behavior to the social environment in which they are situated. They anticipate the reactions of audiences they regard as important (who are also referred to as accountability forums in Bovens’ definition (2007, 2010)), and adjust their behavior accordingly, guided by their motivation to build and preserve a positive image of themselves in the eyes of the relevant audiences (Tetlock, 1992, 1999, 2002). By gaining the approval of their accountability forums, the intuitive politicians seek to avoid negative consequences, to build their own esteem, and to gain power (Tetlock, 1992). These motivational propositions align with the recent research on reputation and accountability in public administration in which the actors’ public image takes center stage (Busuioac & Lodge, 2016, 2017). Following the social contingency model and its many theoretical and empirical extensions, the characteristics of the accountability mechanism, (such as the content of the accountability standard, or the identity of the forum) as well as the context in which it is enacted (such as the type of decision the actor faces), will have a bearing on the decision-making behavior of the accountable actor. Thus, how much effort the accountable actor invests in the decision-making process, or what choice it makes in a dilemma, will depend on the accountability mechanism and the decision context.

In this dissertation, the propositions of the social contingency model are used as theoretical guides in the exploration of the behavioral effects of accountability in the domain of public administration. The application of this model to the domain of public administration is facilitated by the apparent fit of many of its components to the realities of the functioning of the public sector. In addition, the strong empirical support that the propositions of the social contingency model have received

in the behavioral literature (Lerner & Tetlock, 1999), and their corroboration in public administration research (Busuioc & Lodge, 2016, 2017), lend the model credibility. The mechanisms proposed by the social contingency model are carefully applied to the public administration context, and examined experimentally.

1.5. Research questions

This dissertation sets out to advance the understanding of the behavioral effects of accountability mechanisms in the public sector. It specifically focuses on the decision-making behavior displayed by civil servants with some autonomy in situations where they expect to be held accountable by one or multiple forums. It therefore aims to address the following overarching research question:

How do accountability mechanisms shape decision-making behavior in the public sector?

In order to investigate this overarching research question systematically, four narrower sub-questions are defined. Each of them provides the basis for the empirical investigation presented in the four subsequent chapters, and thus guides a sub-project of the dissertation.

Sub-question 1: *What insights can be drawn from the behavioral research on accountability for its study in the domain of public administration?*

The first sub-question focuses on outlining and systematizing the rich existing knowledge regarding the behavioral effects of accountability produced in the adjacent behavioral disciplines. It does so by means of a systematic literature review of the behavioral and experimental accountability literature, which is presented in Chapter 2 of the dissertation. The aim of this sub-

project is to offer methodological, empirical and theoretical insights, which can be used to design tools, develop hypotheses, and provide theoretical arguments to enrich the study of accountability in the domain of public administration.

Sub-question 2: *How does the specificity of the accountability standard affect decision-making behavior in the public sector?*

The second sub-question focuses on the effect of the accountability standard, or *for what* civil servants are being held accountable. Specifically, here it is investigated whether the decision-making behavior of civil servants varies as a result of different degrees of specificity of the accountability standard. This question is investigated in Chapter 3 of the dissertation, by means of a vignette experiment, on one sample of 151 public administration alumni, and on one sample of 88 public administration students.

Sub-question 3: *Which accountability forums get their demands prioritized?*

In the third sub-question, discussed in Chapter 4 of the dissertation, the abilities of individual accountability forums to influence the decisions of civil servants are a central subject of investigation. Here the focus is on the characteristics of the accountability forums who receive priority by civil servants, and thus, on *to whom* account is being given. This question was investigated using a conjoint experiment, administered on a sample of 600 civil servants in the United Kingdom, and a sample of 603 civil servants in the Netherlands.

Sub-question 4: *How does the number of accountability forums and the alignment of their demands affect the decision-making behavior in the public sector?*

Finally, the fourth sub-question investigates the decision-making behavior of civil servants in the context of multiple accountability pressures. It thus aims to capture changes in decision-making behavior which could potentially arise due to the number of forums exerting simultaneous pressures, and the compatibility of accountability demands. This question is examined in Chapter 5 of the dissertation, where the effects of multiple accountability pressures on decision-making behavior are investigated using a vignette experiment on a sample of 270 regulators in the Netherlands.

1.6. Scientific relevance: Integrating behavioral theories and methods

This dissertation seeks to make three contributions to the study of public accountability: to integrate behavioral theories into its study; to develop experimental tools for its empirical investigation; and to test casual effects of accountability mechanisms. Each will be discussed in more detail below.

Integrating behavioral theories into the study of public accountability. In order to advance the understanding of the effects of accountability mechanisms on decision-making behavior in the public sector, this dissertation draws on theoretical models and insights developed in the behavioral sciences (Hall et al., 2017; Lerner & Tetlock, 1999). Thereby, it contributes to the growing behavioral public administration literature (Grimmelikhuijsen et al., 2017). It specifically applies the *social contingency model of judgment and choice* from the field of social-psychology to the context of the public sector (Tetlock et al., 1989; Tetlock, 1992), in order to develop and empirically test hypotheses regarding the effects of accountability mechanisms on decision-making behavior.

The potential contribution of the social contingency model to the understanding of the behavioral effects of accountability in the public sector has been highlighted by a number of scholars (for example Dubnick, 2005; Koch & Wüstemann, 2014; Schillemans, 2016; Han & Perry, 2020ab). However, the model has not been applied to empirically study the behavioral responses to accountability pressures in the public sector prior to the work done in this dissertation. Therefore, this dissertation makes a theoretical contribution to the study of public accountability by applying and testing the social contingency model of judgement and choice to the context of the public sector in conjunction with insights from public administration. In addition, the empirical findings in the dissertation are used to reflect on the appropriateness of the model for investigating accountability in the public sector, on the basis of which, recommendations for further theoretical development are made.

Developing experimental tools to study public accountability. This dissertation also makes a methodological contribution to the study of public accountability. It presents experimental research as a promising avenue for addressing some of the limitations of the existing accountability literature (Koch & Wüstemann, 2014). Accountability mechanisms in the public sector have traditionally been investigated using qualitative methods, and case studies in particular (Brandsma, 2014; Yang, 2012). While these have proved incredibly useful for mapping accountability systems (for example Acar et al., 2008) and understanding extraordinary events (for example Koppell, 2005; Romzek & Dubnick, 1987), they have so far not been used in a way that builds cumulative knowledge of causes and effects of accountability mechanisms. There is, therefore, a lack of understanding of the causal effects of accountability mechanisms in the public sector (Yang, 2012). Experiments are the strongest methodological tool the social sciences have to offer when it comes to testing and establishing causal relations (James et al., 2017). They allow for the controlled

investigation of the effects of specific elements of accountability mechanisms on decision-making behavior. In this dissertation, vignette and conjoint experiments are employed to counterfactually examine the behavioral responses to varying specificity and compatibility of accountability standards, as well as to different numbers and types of accountability forums. These experiments are placed in realistic scenarios, aiming to capture the decision-making context in which civil servants are commonly placed. To achieve realism in the experimental scenarios, and to assure that the problems investigated are indeed relevant in the work of civil servants, two of the three experimental scenarios were jointly developed with practitioners. The experiments developed, therefore present novel methodological tools, aiming to address certain long-standing questions in the understanding of the effects of public accountability.

Besides highlighting the benefits experimental methods provide, this dissertation also discusses the limitations and challenges in using them to study public accountability. It thereby offers a nuanced picture regarding their advantages and drawbacks, as well as specific advice as to how they can be used to advance the understanding of accountability in the public sector. By furthering the development of experimental approaches to the study of public accountability, this dissertation also contributes to the rising wave of quantitative research on the effects of public accountability (for example Bach et al., 2017; Bouwman et al., 2018; Ossege, 2012; Schillemans et al., 2020).

Test causal effects of accountability mechanisms. Finally, and perhaps most importantly, the experimental investigations conducted as part of this dissertation offer an important empirical contribution to the study of public accountability. The results of the experimental investigations point at specific causal effects (as well as lack thereof) of particular characteristics of the accountability mechanism on decision-making behavior of civil servants. Specifically, the dissertation examines and tests three sets of causal effects. Firstly, it examines the effect of the

specificity of the accountability standard on the degree of decision-making effort and the impartiality of decisions, where it finds only weak tendencies. Secondly, it investigates the effects of the characteristics of the accountability forum on the likelihood of the actor to give priority to its demands, and finds that the forum's threat of sanctions is the most important driver of inevitable prioritization decisions. Finally, it examines the decision-making behavior in an environment of multiple accountability pressures, where it tests the causal effect of the number of accountability forums and the compatibility of the accountability demands on the likelihood to employ high and low-effort coping strategies. The examination finds a link between the two factors in the multiple accountabilities environment and the likelihood to use some low-effort coping strategies, although, not entirely in the anticipated direction.

1.7. Societal relevance: Optimizing accountability

Understanding the ways in which accountability mechanisms shape decision-making behavior in the public sector is also of great practical relevance. The knowledge of specific causal links between accountability arrangements and behavioral outcomes would allow for the strategic calibration of accountability mechanisms to achieve particular (performance) goals, and to reflect specific values (Schillemans, 2016; Yang, 2012). Accountability mechanisms could then be “optimized” according to the circumstances at hand and the goals in mind.

The need to hone accountability mechanisms appears especially relevant in light of the numerous reported problems, failures, and unintended consequences of accountability mechanisms in the public sector, including gaming (Bevan & Hood, 2006; Lægreid & Neby, 2016; Terman & Yang, 2016), buck-passing (Mulgan, 2003b), misplacing priorities (Romzek & Dubnick, 1987), and even

organizational paralysis (Koppell, 2005). Even more so, as such deficiencies in the operation of accountability mechanisms are commonly addressed by introducing additional accountability requirements, instead of better-fitted ones (Dubnick, 2005, p. 397). This has only exacerbated the situation, as it has often introduced the problem of accountability overloads (Brennan, 1999; Flinders, 2011; Halachmi, 2014), without necessarily solving the initial issues. Therefore, the knowledge of how specific accountability mechanisms affect decision-making behavior could be useful not only for their optimization to achieve intended goals, but also for anticipating and addressing some of their unintended consequences.

Policymakers might be particularly receptive to this type of research, as the last decade has been marked by increased interest in integrating knowledge of human behavior in public policies and programs. This is evident in the mushrooming of various “nudge” or “behavioral” units in governments and public sector organizations around the globe (Feitsma, 2018). The knowledge of behavioral responses to policies and programs is thus increasingly seen as highly valuable among policymakers for anticipating their effectiveness.

All of the four sub-projects conducted in this dissertation contribute to the production of this practical knowledge. The systematic review of the behavioral and experimental literature of accountability provides a catalogue of effects that accountability mechanisms potentially produce in the public sector. This provides a pool of hypotheses to be explored and tested in the process of designing accountability mechanisms in the public sector, aiming to reach pre-defined goals and to reflect specific values. The three experimental sub-projects offer more concrete insights as to what kind of behavior to expect from specific accountability arrangements in the public sector, highlighting the roles of the accountability standard, the accountability forums, as well as of simultaneous accountability pressures.

1.8. Outline of dissertation

This introductory chapter lays down the rationale of the research approach taken in the dissertation. It also outlines the overarching research question, as well as the sub-questions which provide the basis for each of the empirical chapters. Chapter 2 presents a systematic review of the experimental behavioral literature on accountability, and thus addresses the question as to what we can learn from the behavioral sciences about accountability and decision-making behavior. The following three chapters experimentally investigate the effects of different aspects, or characteristics, of accountability mechanisms in the public sector. Chapter 3 analyzes whether what civil servants are being held accountable for affects their decision-making behavior. Chapter 4 investigates to whose accountability pressures civil servants are most responsive, and thus, which accountability forums get their demands prioritized. Chapter 5 delves into the problem of multiple accountabilities, and investigates the use of coping strategies to deal with the pressures stemming from the simultaneous presence of multiple forums and conflicting demands. Table 1.1 provides an overview of the four empirical chapters of the dissertation. The final chapter 6 concludes by synthesizing the findings from the empirical chapters, in order to provide an answer to the overarching research question. The theoretical and methodological implications of the dissertation are also discussed here, as well as the limitations of the taken approach and recommendations for further research.

Table 1.1 *Overview of the empirical chapters of the dissertation*

Chapter	Question	Method	Sample
(2) Accountability in the behavioral sciences	What insights can be drawn from the behavioral research on accountability for its study in the domain of public administration?	systematic review	266 experiments published in 211 studies
(3) Accountable for what?	How does the specificity of the accountability standard affect decision-making behavior in the public sector?	vignette experiment	Sample 1: 151 public administration alumni in the Netherlands; Sample 2: 88 public administration students in the Netherlands;
(4) Accountable to whom?	Which accountability forums get their demands prioritized?	conjoint experiment	Sample 1: 600 civil servants in the United Kingdom; Sample 2: 603 civil servants in the Netherlands;
(5) Managing multiple accountabilities	How does the number of accountability forums and the alignment of their demands affect the decision-making behavior in the public sector?	vignette experiment	270 regulators in the Netherlands

2. Accountability in the behavioral sciences: Lessons for public administration

2.1. Introduction²

Public sector organizations act upon their delegated powers from the central government to deliver and implement policies. Accountability is crucial to keep them in check (Peters, 2014). Given the importance of accountability mechanisms for the work of public sector organizations, the study of accountability occupies an important place in the field of public administration (Flinders, 2014; Yang, 2012). Accountability, however, is not a settled term in the literature. In this article we define accountability as a mechanism relating an account-giver to an account holder, which should have an impact on the decisions and behaviors of the account-giver (Bovens, 2010).

Recent reviews and discussions of the literature show the wealth of conceptual, theoretical and empirical studies, and the progress that has been made in understanding accountability in the public sector (Bovens et al., 2014; Schillemans, 2013; Yang, 2012). Studies have for instance focused on the various types of public sector accountability (Behn, 2001; Bovens, 2010; Koop, 2014; Romzek, LeRoux, & Blackmar, 2012; Romzek & Dubnick, 1987), and different normative issues related to accountability (Dubnick, 2005; Halachmi, 2002, 2014; Mulgan, 2014). Recent studies also point to some issues and aspects that need further development. Specifically, it has been noted that the public administration literature often focuses on the failures and inefficiencies of accountability mechanisms in the public sector (Flinders, 2014; Schillemans, 2013; Willems & Van Dooren,

² Chapter 2 has been published as a synthesis report article by Marija Aleksovska, Thomas Schillemans and Stephan Grimmelikhuijsen in the *Journal of Behavioral Public Administration*:

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. <https://doi.org/10.30636/jbpa.22.66>

Aleksovska designed the study, collected the data, conducted the data-analysis for both the systematic review as well as the meta-analysis presented in appendix, presented the data, and wrote most of the article. Grimmelikhuijsen and Schillemans contributed with limited writing to help frame the overall argument and to link the findings from the behavioral studies to key issues in public administration literature on accountability, in addition to general supervision and support.

2012; Yang, 2012). Moreover, several scholars have pointed out that the causal chains through which accountability mechanisms operate have not been sufficiently investigated (Brandsma, 2014; Koch & Wüstemann, 2014; Schillemans, 2013; Yang, 2012, 2014). Thus, we still do not have a very good understanding of the causal effects of accountability mechanisms on individual decision-making, behavior and the outcomes of those activities in the public sector (Flinders, 2014; Yang, 2012).

Simultaneously there is a tradition of experimental research in social psychology and other behavioral disciplines investigating how various accountability pressures affect individual decision-making and behavior. Social psychology research in particular, has shown that accountability has a range of effects on human behavior (e.g. Lerner & Tetlock, 1999). Remarkably, this large body of literature has been largely disconnected from the public administration literature on accountability. The aim of this systematic literature review is therefore to show the advantages of integrating behavioral insights and approaches to the study of accountability in public administration, but also to indicate possible shortcomings of these experimental studies. The main issue, that we will discuss in greater detail in the final section of this article, is that such experiments need to be translated and adapted to public sector settings in order to make meaningful inferences about accountability effects (Schillemans, 2016).

We carried out a systematic literature review of experimental studies on accountability to disclose insights from the behavioral sciences to a public administration audience. Our focus is on experimental studies because the effects of accountability on judgements, attitudes and behaviors in the behavioral sciences have been predominantly studied using experimental designs (Hall et al., 2017; Harari & Rudolph, 2017; Lerner & Tetlock, 1999). While there has been a long tradition in experimentation in the behavioral sciences, public administration scholars have less experience

with this method (Bouwman & Grimmelikhuijsen, 2016; James et al., 2017), even with the recent rise in interest in behavioral public administration (Grimmelikhuijsen et al., 2017). Thus, in this review we provide a comprehensive overview of the experimental research on accountability in the behavioral sciences, with the ambition to systematize both substantive and methodological insights on how to advance knowledge of the causal effects of accountability mechanisms in public administration.

It should be noted that this is not the first review article on behavioral accountability research. Lerner and Tetlock's (1999) review has gained the status of a classic reader. However, with nearly two decades since its publication, there is reason to reevaluate the state of the art. Two recent contributions have responded to this increasing need to review the accountability literature. Hall, Frink and Buckley (2017) provide a systematized qualitative summary of the theoretical and empirical advances, focusing specifically on what they call "felt accountability", which relates to the expectation that one will be held accountable for one's actions or decisions. Harari and Rudolph (2017) provide a quantitative synthesis of the findings from experimental studies focusing on rater accountability. This line of research refers to the effects of various types of accountability on the quality of ratings by various professionals, such as teachers (rating tests), traders (rating stock) or estate agents (rating property). Our review provides an important contribution to accountability scholarship. First, it is interdisciplinary, since it covers the experimental research on accountability across all behavioral disciplines. Second, it is systematic and comprehensive in its scope; it is the first comprehensive review covering all experimental studies investigating accountability mechanisms in the behavioral sciences. Thus, it summarizes the findings of 266 experiments in 211 studies from almost five decades of research. Finally, it relates the findings to the public administration literature and draws conclusions for the field of public administration.

We argue that findings from the behavioral sciences should be integrated in the public administration literature by replicating and extending well-established and relevant effects in public sector contexts. Although behavioral studies have found some clearly established positive effects of accountability on decision-making – such as increased decision quality, consistency and cautiousness– experiments can be sensitive to iterations in materials and contexts, therefore replicating and extending well-known experiments in a public sector context is necessary (e.g. Walker et al., 2018). Doing so will provide a greater understanding of the causal effects of accountability mechanisms and their composing elements and contribute to a balanced assessment of accountability effects in the public sector. Before we present our results and research agenda, we develop three challenges for public administration studies on accountability where a behavioral approach may be of value.

2.2. Accountability mechanisms in the public administration literature

The public administration literature on accountability is wide and expansive and has been summarized in a number of recent volumes (Bovens et al., 2014; Dubnick & Frederickson, 2011; Palumbo, 2017) and articles (Schillemans, 2013; Yang, 2012). In the literature, accountability is often understood as a mechanism (Bovens, 2010), although there are different understandings and definitions. Here we focus on accountability as a mechanism aimed to affect the decisions and behaviors of politicians and bureaucrats.

The growing complexity of governments since the 1960s has been one of the driving forces behind the growing scholarly interest in accountability. Modern governance systems consist of multitudes of organizations with differing levels of independence and public character (Verhoest et al., 2012).

The extensions of the chains of delegation in public service delivery, as well as the diversification of the organizations providing public services, especially the expansion of public private partnerships (Reynaers & Grimmelikhuijsen, 2015; Willems & Van Dooren, 2011), have brought the questions of accountability, responsibility and control to the focus of public administration scholars (Lægreid, 2014; Strøm, 2000). Since the organizations entrusted with public service delivery have gained more independence and power from their political patrons, accountability mechanisms have risen in importance, as the central safeguards of the public interest. Accountability mechanisms are intended to ensure the appropriate behavior and quality of work of public service delivery organizations (Schillemans, 2016). Simultaneously, they also provide these organizations with legitimacy, since being accountable is considered a value in itself in democratic societies (Bovens, 2010).

Accountability mechanisms not only serve multiple purposes, but also take multiple shapes and forms. Although they are often spoken of in singular terms, accountability mechanisms have great variation in their designs and characteristics. Scholars distinguish formal from informal accountability (Romzek et al., 2012), hierarchical from horizontal (Schillemans, 2008, 2011), mandatory from voluntary (Koop, 2014), political from administrative (Romzek & Dubnick, 1987), as well as financial, performance and procedural accountability (Behn, 2001). However, while the existence of different types of accountability mechanisms has been recognized in the accountability literature, the consequences of these differences remain largely unexplored (Schillemans, 2013; Yang, 2012). The characteristics of the account holder and the account-giver, the nature of their relationship, as well as the specifics of the accountability standard have been found to have an effect on a wide range of outcomes, as reported in the social psychology literature (Lerner & Tetlock, 1999). This presents a strong indication that the way accountability

mechanisms are set up in the public sector domain might have important consequences on the outcomes they produce. Thus, *it is necessary to stress that accountability takes multiple forms through the variation in the design of its mechanisms, which could potentially have important consequences for the outcomes of the accountability process.*

Despite the importance of accountability mechanisms in theory, their effectiveness and desirable outcomes in practice have been seriously questioned (Bovens et al., 2008). One strand of the public administration literature has argued the existence of accountability “overloads”. This is the argument that (more) accountability is not always good, since it can undermine efficiency and effectiveness (for an overview see Halachmi, 2014). In his commentary, Brennan argues that the existence of multiple simultaneous accountability mechanisms decreases the effectiveness of any one of them (Brennan, 1999). In one of the few empirical investigations of this claim, Koppell (2005) presents a case study showing that the “multiple accountabilities disorder” is a serious threat to organizational effectiveness. Taking this point even further, a number of scholars argue that accountability decreases the capacity of actors to perform, decreases their intrinsic motivation, and diverts their attention from organizational goals to the points of the “accountability checklist”, and thus conclude that accountability and performance are essentially in conflict, referring to this phenomenon as the “accountability paradox” (e.g., Dubnick, 2005; Halachmi, 2002).

In contrast to the accountability overloads literature, another group of scholars has argued that there is a lack of accountability in many domains, and they thus stress the existence of accountability “deficits” (for an overview see Mulgan, 2014). Here the argument is that the complex contemporary systems of governance exhibit numerous unaccountable pockets of power, for instance where policies are implemented by networks (Klijn & Koppenjan, 2014), autonomous agencies (Schillemans & Busuioc, 2014), public-private partnerships (Hasler et al., 2016), or

regulatory experts (Levi-Faur, 2005). Thus, as a number of scholars have already noted (Flinders, 2014; Willems & Van Dooren 2012; Yang, 2012), *a large part of the public administration literature on accountability emphasizes the various failures and problems associated with it and it is more difficult to gauge from existing studies when accountability mechanisms lead to desired outcomes.*

While a considerable number of scholars have not shied away from making strong claims regarding the dysfunctionalities and drawbacks of accountability mechanisms – as Flinders (2014, p. 664) notes: “the narrative surrounding accountability is focused on deficits, problems, and blame games” - the evidence presented in support of these arguments is unsystematic and based upon unrelated case studies (Bovens et al., 2008; Yang, 2012; Schillemans, 2013). Certain claims have already become “self-evident truths”, since they have been accepted as valid, even in the absence of hard evidence to support them (Flinders 2014, p. 663). This can be in part attributed to the dominant approach used in the empirical analysis of accountability mechanisms in the public sector. Specifically, the empirical investigations of accountability mechanisms in the realm of public administration have been marked with a lack of methodological diversity, with qualitative case studies being the method of choice (Brandsma, 2014; Koch & Wüstemann, 2014; Yang, 2012; Yang, 2014; Schillemans, 2013).

Qualitative designs, and case studies in particular, are quite suitable for exploring and describing phenomena which are nascent in the literature, or developing theories. Case studies can help to study the combined effects of various institutional mechanisms in specific administrative contexts. They can be fruitfully combined with other forms of research and can contribute to the accumulation of knowledge. This, however, presupposes that the researchers build on earlier studies and findings and use similar conceptual and theoretical lenses. Earlier reviews of

accountability studies in public administration have shown that this is often not the case (Yang, 2012; Yang, 2014; Schillemans, 2013). The rich body of accountability case studies is often unrelated and – while scholars may be using the same word (“accountability”) – they may be referring to different definitions and, thus, phenomena. In this context, experimental research can add a lot of value to our knowledge of accountability in public administration. Most experimental studies of accountability in the behavioral sciences are based on a shared definition of accountability which facilitates the accumulation of knowledge (Lerner & Tetlock, 1999). This research could serve as the basis for the development of a public administration experimental research program on accountability. By building on an already developed cumulative body of literature, as well as using the advantages of methodological rigor that experimental designs possess, experiments could help advance the knowledge of causal effects of accountability mechanisms in the public sector (Yang, 2012; Koch & Wüstemann, 2014). In sum, *while causal effects of accountability mechanisms are often discussed, the public administration literature has so far not explored and rigorously tested those*. Experimental research can add much value to the existing body of knowledge on public sector accountability.

In summary, we identify three challenges in the current literature on accountability mechanisms in the public sector which can be informed by existing experimental studies:

- a) The challenge of establishing causal effects of accountability mechanisms on individual decisions and behaviors,
- b) The challenge of establishing a balance between desirable and undesirable effects of accountability, and
- c) The challenge of accounting for the variation of characteristics of accountability mechanisms and its effects.

By looking at the rich experimental literature on accountability produced in the behavioral sciences, we aim to provide insights and directions towards overcoming these issues.

2.3. Systematic review: Methods and data

The research reviewed in this article has been selected using a systematic process, as outlined in the PRISMA guidelines (Tranfield et al., 2003). In particular, the search terms “accountability” and “experiment” were used to identify the relevant studies, in the entire corpus of books and articles in the behavioral sciences listed in five major databases of academic literature, namely, Web of Science, JSTOR, EBSCO, Scopus, and Science Direct.³ This search yielded in total of 7022 results. These results were complimented with an additional 344, which comprise the complete reference lists of three review articles on accountability research.⁴ The inclusion of the reference lists of previous review articles was done with the purpose of cross-validating the results of the database search, as well as capturing the studies which do not use the keywords used in our database search, but refer to the concepts of interest.

After the removal of duplicates, a total of 6559 unique documents were identified for preliminary screening. As a second step in the selection process of the studies reviewed in this article, the abstracts and often the methods sections of the studies were screened to determine that at minimum the study in question is an experimental study that researches the concept of accountability in its broadest sense. This screening resulted in the selection of 594 articles.⁵ These articles were then closely examined in order to determine whether they meet the two selection criteria described

³ Complete search log and a PRISMA flow diagram are available in appendix.

⁴ These articles are Lerner and Tetlock (1999); Hall et al. (2017) and Harari and Rudolph (2017).

⁵ Only one relevant book chapter was found, as well as three dissertations.

below. 211 articles, presenting 266 unique experiments satisfied the two criteria and formed the basis for our review.

Selection criteria

The selection of the articles was based on two criteria which reflect the focus of this review, and our understanding of the two central concepts guiding this study: accountability and experiment.

Accountability. In this study the concept of accountability is understood in its narrow sense (Bovens, 2010). Specifically, we only included studies that treat accountability as a relationship between an actor and a forum, in which the actor has an obligation to justify his or her conduct (Bovens, 2010). The relationship between an accountability forum and an actor can take the form of a principle-agent relationship, with the actor being the agent, and the forum being the principle. However, the forum is not necessarily always a principle. The forum's authority to demand justifications for the actor's behavior can come from sources other than delegation, such as for example professionalism and expertise (Bovens, 2010). Studies which treat accountability as mere presence of another person, or identifiability of the actor are not included in this review. The accountability mechanisms in the public sector are characterized by an expectation of evaluation and need for justification, and we chose to select studies that treat accountability only as such. Thus, the first criterion of inclusion is that the concept of accountability is treated as a mechanism which defines the relationship between an actor, and a forum that holds the actor accountable.

Experiment. The second crucial criterion for inclusion is the usage of the dominant method in accountability studies – the experimental method. Here we rely on a fairly universal definition of experiments, which states that for a study to be considered an experiment, the researcher must intervene in the data generating process by consciously manipulating elements of it (Morton &

Williams, 2010, p. 42). Thus, we included ‘laboratory experiments’, ‘survey experiments’, ‘field experiments’ and ‘lab-in-the-field experiments’ in our analysis. Natural experiments were excluded, as these do not involve a manipulation by the researcher. They are also strongly dependent on specific contextual circumstances which also affects accountability, which makes this type of experimentation, although highly relevant, difficult to compare to the more controlled experiments.

All of the included experiments involve a comparison between treatment groups. The allocation of participants in treatment groups has been performed using a randomization procedure in the majority of the studies, with the exception of the experiments with repeated measurement designs.

Overview of the data

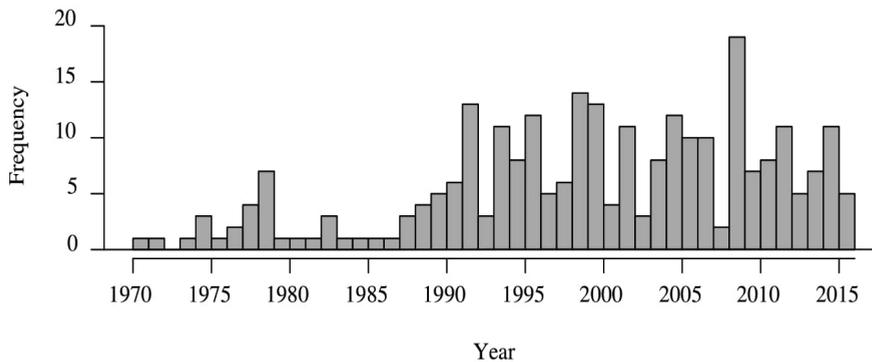
The 211 selected experimental studies span the period between 1971 and 2016.⁶ The earliest experiment analyzing the effects of accountability mechanisms dates back to 1971, and is followed by a steady rise in interest in this topic over the decades (Figure 2.1). The largest proportion of experiments is conducted within the realm of psychology (40%), followed by organizational sciences (16%) and accounting (11%). These studies appear in a wide range of outlets, specifically 96 different journals, three dissertations and one book chapter.

The experiments themselves are quite uniform: the overwhelming majority of them are laboratory studies (83%), with between-subjects design (83%), and students as experimental subjects (73%). Most of these experiments are conducted in the US (54%), with the second highest – the Netherlands – having significantly fewer (7%). In fact, less than 5% of the studies are conducted

⁶ The complete list of studies included in this review is available in Appendix.

in countries which are not western developed democracies. But while these studies come from a variety of disciplines, they seem to be performed in a fairly similar setup: in the classrooms of campuses of western universities.

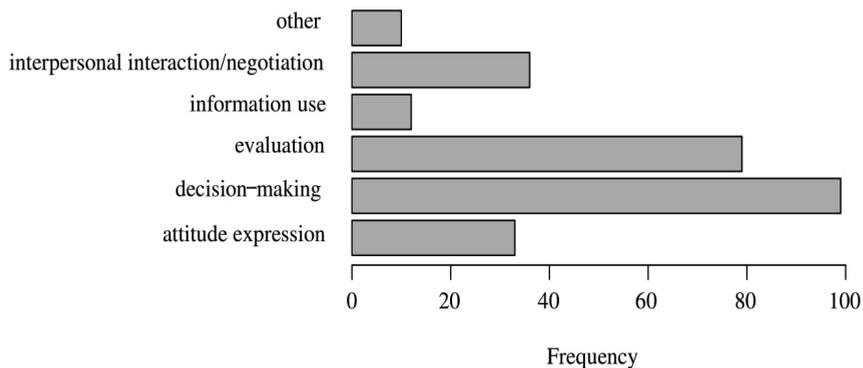
Figure 2.1 *Frequency of conducted experiments over time*



Concerning the substantive design of the reviewed experiments, the participants have been placed in a variety of settings with different degrees of realism. The setting describes the context in which the accountability mechanism has been investigated and thus provides more detailed information regarding the setup of the experiment. Specifically, the most frequently observed settings are audit evaluations (9%) and negotiations (9%), followed by performance evaluations (5%), and attitude expressions (3%). Similarly, we have also recorded the type of task the participants were asked to perform to have a better insight as to what the experimental participants were expected to do, and what they were held accountable for. As Figure 2.2 shows, decision-making and evaluation tasks dominate the experimental designs in the study of accountability mechanisms. A closer look at the

results of these studies reveals what kinds of research goals these experimental tasks have been set up with.

Figure 2.2 *Type of experimental task frequency*



2.4. Effects of accountability mechanisms in behavioral research

The 266 selected experiments examine a wide variety of relationships and effects. In order to facilitate their discussion, we organized the findings in four main themes: 1) decision-making, 2) behavior, 3) external outcomes and 4) characteristics of the accountability mechanism. The logic of the chosen organization is explained below.

Since these experiments treat accountability as a mechanism, as opposed to observed behavior or democratic value (Bovens, 2010), in all of them accountability appears as an independent variable. However, not all of the experiments take the same approach in investigating accountability mechanisms. Specifically, one group of experiments examines the effects of accountability

mechanisms by contrasting the dependent variable of interest when such mechanism is present and when it is not. Thus, in these experiments one group of participants is subjected to an accountability mechanism, while another group is not, and the effect of the accountability mechanism is established by comparing the results of the two groups. A second group of experiments, in contrast, compares the effects of two (or more) accountability mechanisms, which differ in a particular characteristic. In these experiments all groups of experimental participants are subjected to an accountability mechanism, however, the characteristics of the accountability mechanisms differ across groups. By comparing the results of the experimental groups, these studies show how the characteristic of the accountability mechanism can affect a certain outcome of interest.

The results of the first group of experiments – comparing the results when an accountability mechanism is present and when it is absent – are organized in three themes: decision-making, behavior, and external outcomes. These themes were identified with the following procedure. First, we recorded all tested main effects of accountability mechanisms on the various outcomes explored in these studies, and the evidence the conducted experiments provide in support of their existence. We then clustered the tested effects based on their similarity. The results of the second group of experiments – comparing the results of two or more accountability mechanisms with different characteristics – are considerably smaller in number, and thus presented as a separate theme.

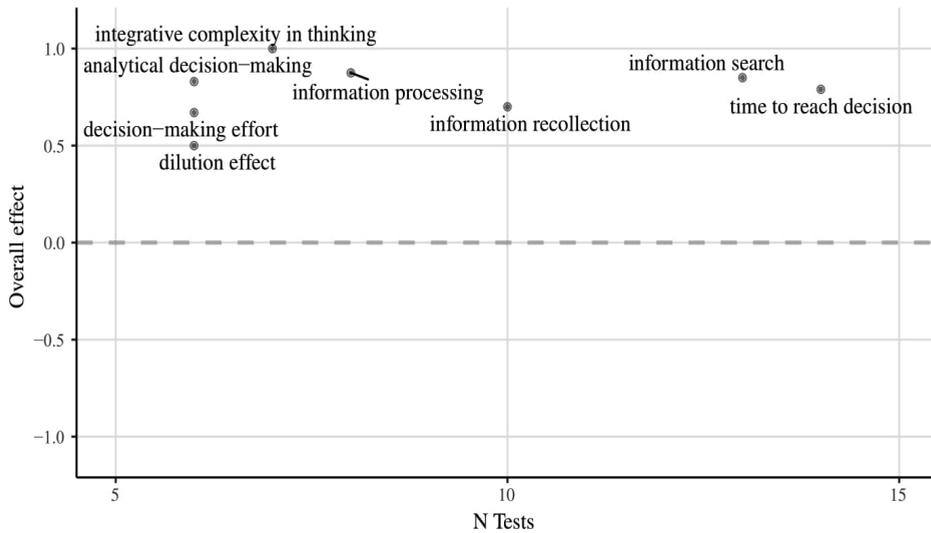
In this review, we present the examined relationships of accountability mechanisms and its characteristics on a wide range of effects, and we report whether or not the researchers found support for these relationships.

Accountability and decision-making

As decision-making is at the heart of public administration (Simon, 1947), the effects of accountability on decision-making are highly relevant for our purposes. The behavioral sciences have delved into the study of decision-making, its components, characteristics and biases, and thus offer us an important resource to better understand decision-making in the realm of public administration.

The way in which the presence of an accountability mechanism affects the different components of the decision-making process is perhaps the most central theme explored in the behavioral and experimental literature on accountability. Figure 2.3 summarizes the central findings of this literature, by presenting the effects that have been tested in at least five different experiments.⁷ The x-axis presents the number of studies that have examined the relationship between an accountability mechanism and the particular dependent variable, while the y-axis presents what we call an overall effect. The overall effect presents a summary of the findings on the studied relationship. The extreme scores of 1 and -1 denote that all studies examining the particular relationship have found a positive, for 1, or negative for -1, effect. In the space between the extreme scores 1 and -1 are the studied relationships for which the studies do not provide consistent effects, either because of observed null effects, or effects pointing in different directions (this applies to both figures 2.4 and 2.5 as well).

⁷ A detailed representation of all effects is provided in appendix, organized by theme. In addition, they are also available in the form of a searchable online database: <https://aleksovska.shinyapps.io/accountability/>

Figure 2.3 *Effects of accountability on decision-making*

Note. The x-axis represents the number of times the effect has been tested experimentally, while the y-axis presents the direction of the effect (positive effects appear on the upper part of the graph, while negative at the bottom part of the graph), as well as the robustness of the effects (the placement of the variable on the y-axis is a proportion of the sum of the effects across studies with 1 being positive, -1 negative, and 0 no effect, and the number of studies, for example, if an effect has been tested five times and three times found positive, and two times no effect was found, the effect is calculated as $(3+0)/5=0.6$).

The experimental studies investigating the effect of accountability on decision-making consistently found that decision-makers facing accountability, searched for (Huneke et al., 2004; Lee et al., 1999)⁸ and remembered (De Dreu et al., 2006; Roch, 2006; Tetlock, 1983b) more information than decision-makers not facing accountability. They also engaged in deeper information processing (Hess et al., 2001; Thompson et al., 1994), used more analytical decision-making strategies (Huneke et al., 2004), and displayed higher integrative complexity in thinking than decision-makers who did not face accountability demands (Green et al., 2000; Tetlock et al.,

⁸ Due to space limitation issues, not all relevant articles are referenced in the discussion of the findings. A full list of the articles can be found in appendix.

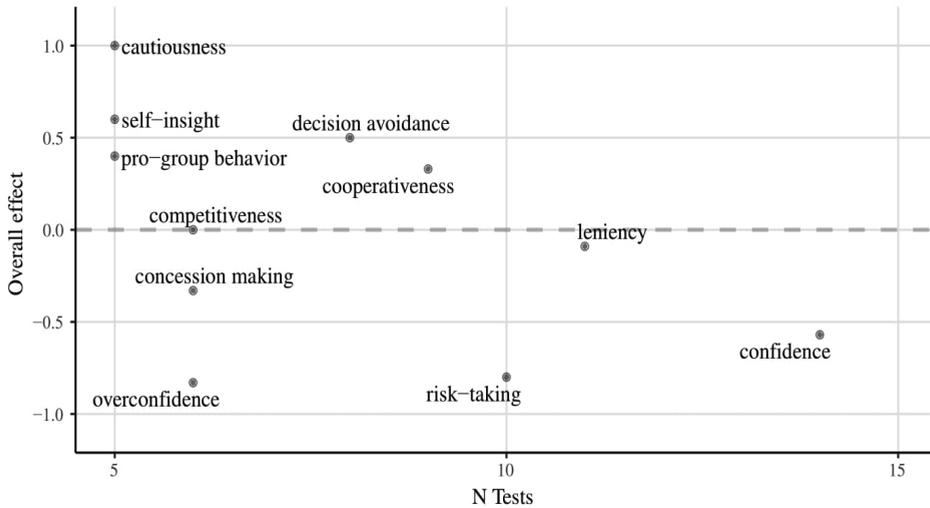
1989). Thus, it is hardly surprising that decision-makers facing accountability have been found to invest more effort in the decision-making process (Frink & Ferris, 1999), and to take more time to reach their decisions (Van Hiel & Schittekatte, 1998; Lee et al., 1999) than ones not facing accountability pressures.

The effects of accountability have also been tested on a number of decision-making biases. A large number of these effects have however only been investigated in a single, or couple of studies, and thus they have not been validated across multiple settings. Nevertheless, these studies show that in the case of most decision-making biases, accountability has an attenuating effect. A notable exception here is the dilution effect, which is the tendency of people to underutilize relevant information when irrelevant information is also present. This judgement bias has been found to be enhanced by accountability in several studies (Tetlock & Boettger, 1989).

What these findings offer as a general conclusion is that accountability has an overwhelmingly positive effect on decision-making. Specifically, accountability improves the collection and treatment of information and stimulates more effortful decision-making. These results should serve as a reminder of the benefits of having accountability, and a counterbalance to the critical discussions of accountability.

Accountability and behavior

The accountability mechanisms in the public sector are introduced with the intention to keep the behavior of public servants in check. Thus, there is an underlying assumption that the pressure of accountability does affect behavior in predictable and desirable ways. Whether this is indeed the case has been explored extensively in the behavioral literature on accountability. The main findings of this research are presented in Figure 2.4.

Figure 2.4 *Effects of accountability on behavior*

Perhaps not surprisingly, the presence of accountability pressure has been found to stimulate a great deal of defensive behavior: individuals who are held accountable have been found to avoid taking risks (Weigold & Schlenker, 1991) and decisions more generally (Tetlock & Boettger, 1994). These behaviors can easily be associated with the negative effect that accountability has been found to have on confidence (Tetlock & Boettger, 1989; Thompson, 1995) and overconfidence (Jermias, 2006; Tetlock & Kim, 1987), as well as, conversely, the positive effect on self-insight (Johnson & Kaplan, 1991) and cautiousness (Ready & Young, 2015). While the pressure of accountability might hinder the decision-making process, it can also stimulate cooperation (Coletti et al., 2005) and motivate actions which enhance the wellbeing of groups (Barreto & Ellemers, 2000).

The effects of accountability on competitiveness and leniency⁹ in our summary of the results are zero or close to zero. In reality, while many of the tested effects on these behaviors are null, the ones which are not point in different directions and therefore they cancel each other out. This indicates that the effect of accountability on these behaviors depends on the context in which they are placed.

In sum, the findings presented here show that accountability stimulates socially-conscious, cautious, and risk-averse behaviors, which is what we would like to expect from a responsible public sector. However, these virtues can turn into failures when they appear in excess, and block the decision-making processes. Under which conditions accountability stimulates responsible decision-making and under which it results in a decision-making deadlock and decision avoidance is not yet clear. Such knowledge could be of great value in the design on accountability mechanisms in the public sector.

Accountability and outcomes

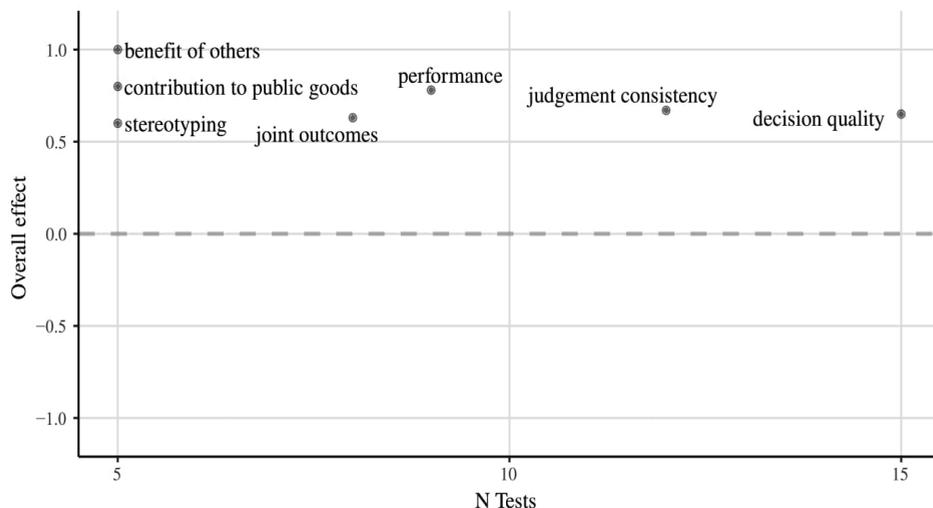
Bovens and colleagues (2008) have argued that accountability in general serves three goals or outcomes in public administration. These three understandings of accountability are: first, accountability as a tool for democratic control; second, accountability as a safeguard of the rule of law; and third, accountability as a tool for learning and improving policies. In experimental studies, the effect of accountability has also been investigated on a myriad of outcomes produced by decision-making and behavior under accountability pressures, although these outcomes are defined on much lower levels of abstraction. These outcomes are treated as a separate category

⁹ Leniency is understood here as the tendency to make decisions which are favorable to the person who is directly affected by that decision. It is estimated as a difference in favorability in two of more conditions, for example, in a situation where an evaluator needs to justify his evaluation in front of a superior, or to the person being evaluated, or does not need to justify his evaluation at all.

here since they do not present an integral part of the decision-making process of behavior, but represent results of those processes, usually linked to particular contexts. Figure 2.5 presents the most researched, and thus most important effects in the behavioral experimental literature.

The results from experimental investigations indicate that accountability enhances decision quality (De Dreu et al., 2000; Scholten et al., 2007; Schul & Mayo, 2003), improves judgement consistency (Ashton, 1992; Hagafors & Brehmer, 1983; Ordóñez et al., 1999) and task performance (Frink & Ferris, 1999; Holt et al., 2012; Tan & Kao, 1999). Accountability also produces positive effects in the social domain, as the results indicate. In particular, accountability pressures have been found lead to greater benefit to others (Schoenrade et al., 1986), greater contributions to the public good (Lierl, 2016), as well as to greater joint outcomes in negotiated agreements (De Dreu et al., 2006).

Thus, these findings consistently show that accountability promotes decisions and performance of higher quality, and a number of desirable social outcomes. They thus stand in sharp contrast to the dominant negative narrative found in the public administration literature on accountability.

Figure 2.5 *Effects of accountability on decision-making and behavior outcomes*

Note. Decision quality has actually been experimentally tested 26 times, but for ease of representation it has been placed at the maximum on the displayed scale.

Characteristics of accountability mechanisms

In the public administration literature, authors have clarified that there are many, often quite divergent, types of accountability (e.g. Bovens et al., 2014). When talking about the effects of accountability mechanisms in public administration, their specificities and characteristics are often left out. One sizeable part of the experimental research on accountability in the behavioral sciences has investigated the elements of accountability mechanisms and their effects on a wide range of outcomes. Table 2.1 presents an overview of the characteristics of accountability mechanisms which were compared against each other in this literature, and the scholarly attention these contrasting pairs have received. Following Schillemans' CPA-model (2016), we organized them in three thematic groups: relating to accountability timing, to the accountability standard, or to the relationship between the actor and accountability forum.

Table 2.1 *Studied characteristics of accountability mechanisms*

Contrasts of accountability mechanisms characteristics	Number of experiments	Number of tested effects
<i>Timing</i>		
Pre- vs. post-information accountability	6	8
Pre- vs. post-decisional accountability	5	8
Expected vs. unexpected accountability	1	1
Continuous vs. end accountability	1	1
<i>Standard</i>		
Process vs. outcome accountability	14	24
Forum with known vs. unknown views	4	7
Forum with homogeneous vs. heterogeneous views	2	3
Forum with conflicting vs. supporting views	1	1
Forum with rule conflicting vs. supporting views	1	1
Aggressive vs. conservative forum	1	1
<i>Relationship</i>		
Forum as authority vs. beneficiary	7	16
Individual vs. group accountability	5	17
High vs. low forum standing	5	5
Expected vs. no expected forum meeting	4	4
Accountability to ingroup vs. outgroup	1	4
Accountability for decisions affecting others vs. affecting only self	3	3
<i>self</i>		
Actor with high vs. low standing	2	2
Small vs. large size form	1	2
External vs. internal accountability form	1	1
Identifiable vs. unidentifiable actor	1	1
Top-down vs. top-down and bottom up accountability	1	1

As Table 2.1 shows, the effects of timing have been the least researched characteristics of accountability mechanisms. Despite the small size of this research theme, several effects have consistently been found. In particular, expecting to be held accountable for the decision prior to making it, as opposed to being held accountable for it afterwards, has been found to increase decision quality (Schul & Mayo, 2003), and integrative complexity in thinking (Tetlock et al.,

1989). Setting the accountability expectation before, as opposed to after the task information has been provided, has been found to reduce decision-making biases such as primacy effects (Tetlock, 1983b), recency effects (Kennedy, 1993), and attribution errors (Tetlock, 1985). In sum, these findings show that setting an accountability expectation before any element of the task is presented, results in less biased decisions.

The accountability standard, or what an actor is being held accountable for, has received considerable interest from behavioral researchers. The dichotomy of accountability for process and accountability for outcome is central in this research theme, and has produced a number of findings with high potential relevance to public administration. Specifically, while process accountability has been found to be better at producing quality decisions which are thought through, calibrated and consistent (Scholten et al., 2007), outcome accountability has been found to result in greater commitment to the decisions made (Simonson & Staw, 1992). In addition, process accountability has been found to stimulate performance improvement, while outcome accountability has been found to yield better performance (Davis et al., 2007).

The rest of the studies that have looked at accountability standards have explored how the views and expectations held by the accountability forum affect the decision-making and behavior of actors facing accountability pressure. These studies point out that the knowledge of the forum's views has more drawbacks than advantages, stemming from the actor's motivation to satisfy the forum. Thus, the actor who is aware of the forum's preferences has been found to display attitude shifts (Tetlock, 1983a), and to put less effort into collecting and analyzing information relevant to the decision-making process (Tetlock et al., 1989).

The third group of studies investigating the characteristics of accountability mechanisms dissects the relationship between actor and accountability forum. Several aspects of this relationship have been explored, including the size of both the forum and the actor, their standing, as well as the position they hold in relation to each other. This research finds that when the forum is in a position of authority, as opposed to being the beneficiary or the receiver of the decision, actors tend to feel more responsible for their decisions, they are harsher in their evaluations and less lenient in their decision-making, and essentially produce decisions of higher quality (Mero et al., 2007; Pennington & Schlenker, 1999). Decisions of higher quality have also been found to be achieved when the forum has a high standing, as opposed to low (Sedikides et al., 2002), and when the forum is considered to be ingroup, as opposed to outgroup¹⁰ (Haccoun & Klimoski, 1975), as well as when the actor is individual, as opposed to a group (Kroon et al., 1991).

In sum, these findings show that the characteristics of an accountability mechanism can have a great impact on the effects produced by that very accountability mechanism. This point has been to a large extent neglected in the public administration literature on accountability. Our review provides initial insights into how the characteristics of the accountability mechanism influence decision-making and behavior. However, these findings have not been tested in the context of public organizations. Since the public sector is increasingly subject to all kinds of accountability pressures, we need more research to assess whether and how these characteristics hold in the field of public administration.

¹⁰ The distinction between ingroup and outgroup is commonly used in psychology to contrast groups with which the person identifies as a member or not. In the study reviewed here the decision-maker facing accountability pressure was either to give account to members of its own group, or to strangers.

2.5. Research agenda for accountability in public administration

The analysis in this study shows that accountability has been found to have far reaching effects on individual decision-making, behavior, as well as for the outcomes of that behavior. Almost five decades of experimental investigation on accountability have resulted in a list of well-established effects with regards to the relationship between accountability and decision-making depth and quality, defensive and cooperative behavior, as well as to the social considerations of the behaviors performed under accountability pressures. Many of these established effects have direct relevance to public administration scholarship. However, for now, accountability research in public administration and the behavioral sciences exist in parallel – not in syntheses.

However, simply applying experimental findings to the public sector would be a mistake. We observed a few potential shortcomings in the behavioral literature. First, the experiments we analyzed for this article were mostly carried out in a stylized context, such as laboratories, which is not directly related to the public sector. Second, 73% percent of these experiments were carried out with student samples, which might limit generalizability to the public sector, because students are not commonly exposed to accountability pressures in the way a seasoned public manager is. Overall, we simply do not know if there is ‘something’ in public organizations that causes individuals to respond differently to accountability mechanisms (Schillemans, 2016) and we should therefore translate and replicate behavioral and psychological insights in public sector contexts to bridge the gap with public administration scholarship on accountability (Grimmelikhuijsen et al., 2017; Walker et al., 2018). To do so, we propose three points for future research to integrate methodological and substantive insights from the behavioral sciences into public administration.

1. Replicate and extend well-established effects in public sector contexts

Accountability scholars in public administration have pursued a less focused research agenda than their colleagues in the behavioral sciences with more diverging conceptualizations, definitions and measurements. As a result, the theorized causal links between public sector accountability mechanisms and the stated (un)desirable outcomes have not been tested rigorously (Brandsma, 2014; Koch & Wüstemann, 2014; Yang, 2014). Often when behavioral consequences of accountability are explicitly mentioned, scholars primarily maintain a theoretical discussion or explore effects using qualitative research designs. Based on our review we have three recommendations for future research.

First, public administration can use existing methodological tools to test causal relationships. Specifically, the long tradition of experimentation with accountability in the behavioral sciences allows public administration scholars to apply tried and tested experimental scenarios, designs and manipulations. These tools can serve as a learning tool, helping to develop public administration's own experimental perspective on accountability (cf. James et al., 2017). In addition, the findings in experimental studies of accountability can be combined, provided researchers are strict and precise in their use of theories and concepts, with insights generated in surveys or with qualitative research designs, in order to accumulate knowledge.

Second, this article has systematically analyzed the experimental evidence on the effects of accountability from a public administration perspective, guided by the main concerns and issues expressed in that literature. This has led to a discussion of main findings, issues, conclusions and also discrepancies between experimental research in psychology and other disciplines with public administration research. We have thus adopted an outside-in perspective on the experimental

studies. As a further step, it would make sense to reverse and also apply an inside-out perspective to gauge effect sizes, recurring findings but also areas of uncertainty. In short, to further the analysis of experimental evidence in a meta-analysis. A meta-analysis integrates comparable findings from various studies and helps to establish amongst others effect sizes and robustness of findings (Field & Gillett, 2010; Hunter & Schmidt, 2004). We have taken a step in that direction, by meta-analytically analyzing six effects of accountability on decision-making in the appendix. A thorough approach, however, in assessing the effect sizes and robustness of the findings discussed here would be helpful to further set the agenda of experimental studies of accountability.

Thirdly, behavioral studies also provide a range of well-established findings, yet these experiments have often been performed in laboratory settings, with neutral and low-stakes scenarios. It is therefore relevant to translate these findings to more realistic settings in the public sector, taking into consideration the professional training and experience of public sector employees, managers and political actors, as well as the nature of their tasks. Scholars could start off by replicating and extending the most well-established and relevant findings from laboratory settings to real public administration settings. For instance, there are many experimental studies demonstrating the positive effects of accountability on the quality of decision-making in the lab in terms of integrative complexity of reasoning. Under conditions of accountability, participants process and integrate more information and take more thoroughly considered decisions. A fruitful research question would be whether we can see the same effects in real world settings of complex decision making, for instance by public managers or regulators. For behavioral effects, public administration scholars should be interested in experiments showing increased cooperation but also decision avoidance. Finally, and perhaps of greatest importance for the public sector are the large number of experiments showing positive effects of accountability on decision quality and consistency.

2. Towards a more balanced assessment of accountability effects

The effectiveness and desirable outcomes of accountability mechanisms in practice have been seriously questioned (Bovens et al., 2008). For instance, more accountability can undermine efficiency and organizational effectiveness (Halachmi, 2002; Koppell, 2005). The often critical findings in the public administration literature stand in contrast with the studies in our review. Most behavioral experiments indicate far more positive and desirable effects of accountability than what the public administration literature argues and finds. This may be affected by a publication bias but is more strongly related to research designs. Many experimental studies relate normatively positive dependent variables to accountability as either an independent variable or a moderator, as will be discussed below. For instance, various experiments have shown that accountability – in certain forms – leads to improved decision quality (De Dreu et al., 2000; Schul & Mayo, 2003) or better cooperation between actors (Coletti et al., 2005).

This contrast with the findings from the public administration literature on accountability on the one hand, and the behavioral literature on the other. This could be due to several reasons. First, accountability as a theme in the public administration literature has been said to be a “critical companion to administrative change” (Schillemans, 2013). Thus, it has often been discussed as a problematic issue arising from new administrative reforms, such as service delivery in public-private partnerships or the delegation of tasks to international (EU) or sub-national (agencies, local government) levels. Secondly, the issue of accountability is often retrospectively invoked in studies of systemic crises or policy failures (e.g., Romzek & Dubnick, 1987; Koppell, 2005). Analyzing what went wrong in a particular situation in the public sector then easily spills over into critical assessments of accountability in a given situation. As the point of departure of such studies is a given crisis or failure which is often analyzed as a rich (and valuable) case study, there is no

counterfactual. Thirdly, in many critical public administration studies, scholars point at the excessive costs of accountability in terms of for instance red tape, opportunity costs or its alienating effects on public sector professionals and street-level-bureaucrats (Halachmi, 2014; Behn, 2001; Bovens et al., 2008). These studies in essence focus on important negative side-effects of accountability mechanisms. They show that actors may find accountability a nuisance or that it comes at unacceptable costs. Behavioral studies have in contrast focused more strongly on the main, normatively positive, effects of accountability, such as the quality of decision-making or compliance with specific norms. These different types of studies are in principle not mutually exclusive at all. Accountability mechanisms may indeed have the desired main effects (such as compliance) as suggested by behavioral scholars yet they may come at the cost of severe side effects (such as demotivation or unacceptably high costs) as suggested by public administration scholars.

All in all, it seems that at least quite some of the discrepancy in the tone of findings between psychological and public administration research on accountability is the product of a different focus in research. Compared to psychologists, public administration scholars have focused much more on institutional reforms challenging accountability, on cases of failure and crisis, and on some important negative side effects of accountability. Behavioral scholars on the other hand have put more focused attention on the desired effects of accountability mechanisms and their causal relation to individual behaviors and decisions. The more positive findings of the behavioral literature on accountability are potentially relevant for public sector accountability, although we should further test to assess to what extent experimental findings from the lab are moderated in the field of public administration and interact with contextual factors (Schillemans, 2016). By developing experiments that test both potential positive and negative effects of accountability on

decision-making and decision outcomes, with clear counterfactuals, public administration scholarship develops a more balanced understanding of accountability effects. And experimental knowledge on the effects of accountability could be integrated with more traditional qualitative and quantitative approaches in the field.

3. Focus on the multiple design elements of accountability

Third, and final, the findings of this study show that the behavioral sciences have investigated a number of specific characteristics of accountability mechanisms, relating to the timing of accountability, the evaluative standard, as well as the relationship between the actor and the forum, which show that the way the accountability mechanism is constructed has important consequences too. While public administration scholars have recognized the existence of great diversity in the types and characteristics of accountability mechanisms, little attention has been allocated to investigating the consequences of this diversity. Thus, what has not received sufficient attention so far is the fact that *how* accountability mechanisms are designed could strongly affect the behavioral and decisional outcomes.

For instance, we have found that the content of the accountability standard (process or outcome) (Scholten et al., 2007; Simonson & Staw, 1992), the relationship of the accountability giver and the accountability holder (internal or external) (Haccoun & Klimoski, 1975), can encourage the account-giver to think and act in different ways. The public administration literature has discussed problems of accountability multiplicity and overload (Koppell, 2005), however we still have little understanding as to how multiple accountability mechanisms with similar or different characteristics interact with each other in the public sector, and bring about some of the observed consequences. Thus, it is necessary to move beyond the understanding of accountability as a

unitary phenomenon, and to look more closely at the design elements of which the accountability mechanisms consist, how they interact, and how they affect the outcomes of interest. This is of utmost importance for the study and practice of public administration. Since accountability is a constant and pervasive condition in the public sector, what can and should be modeled is the design of the operating accountability mechanisms.

3. Accountable for what? The effects of specifying the accountability standard

3.1. Introduction¹¹

Accountability mechanisms are universally present in the public sector. They are relationships between an actor and a forum, in which the actor has an obligation to justify his or her conduct to the forum (Bovens, 2007; 2010). Their characteristics, however, tend to vary greatly. Scholars usually distinguish four core elements of any accountability relationship, summarized in four short questions: “Who?”, “To whom?”, “For what?” and “Why?” (Bovens, 2007; 2010; Mulgan, 2003a). These four questions relate to the identity of the account-giver or the actor, the identity of the account-holder or the forum, the accountability standard, and the nature of the relationship between the two, accordingly. Each of these four questions can receive a number of different answers when describing the accountability relationships in the public sector. Thus, public sector accountability mechanisms can take multiple different forms. While these variations in the characteristics of accountability mechanisms in the public sector are well-recognized (for example Behn, 2001; Koop, 2014; Romzek & Dubnick, 1987; Romzek et al., 2012; Schillemans, 2010), their consequences seem insufficiently investigated and understood (Aleksavska et al., 2019; Bovens, 2010; Dubnick, 2005; Schillemans, 2016; Yang, 2012). The present state of the art of the public administration accountability literature does not give us a clear indication whether or how one’s decision-making behavior is shaped as a result of these variations in the characteristics of the accountability relationships, nor what the potential consequences of those behaviors might be.

¹¹ Chapter 3 has been published as an original research article by Marija Aleksavska in *Public Performance & Management Review*.

Marija Aleksavska (2021) *Accountable for What? The Effect of Accountability Standard Specification on Decision-Making Behavior in the Public Sector*, *Public Performance & Management Review*, <https://doi.org/10.1080/15309576.2021.1900880>

Understanding the effects of accountability mechanisms and their varying characteristics, are, however, of great practical value for policymakers and practitioners. Their strategic calibration could help shape individual behavior in predictable ways, and thus aid in the achievement of performance goals (Schillemans, 2016; Yang, 2012). Conversely, their misalignment with organizational goals and values could lead to less than optimal outcomes, and even problems and failures (Overman, 2020; Romzek & Dubnick, 1987; Terman & Yang, 2016). Accountability mechanisms are, thus, potentially very important management tools, whose powers remain insufficiently explored (Dubnick, 2005).

In the last few years, policy-makers in a few European countries have opened a discussion about the need for calibration of the *accountability standard*. The starting point of this discussion is the perception that the professional environment of (some) civil servants is over-saturated with rules and procedures, for which they are held accountable (Brennan, 1999; Halachmi, 2014; Pollitt, 2003, p. 47; Power, 1997, 2003). It has been argued that these stringent accountability requirements impose significant compliance costs, and thus decrease performance efficiency (Dubnick, 2005; Jos & Tompkins, 2004; Warren, 2014, p. 44). Additionally, they are also found to “rob” civil servants of their professional pride, since they signal mistrust, and reduce their professional autonomy (Harrison & Dowswell, 2002; Hoecht, 2006; Power, 1997; Schillemans, 2016; Warren, 2014, p. 44). The requirement to comply with strict accountability rules and procedures has been also argued to introduce rigidity in decision-making processes and reduce the ability of civil servants to tailor their responses to the characteristics of the problems presented (Daly, 2009; Molander et al., 2012; Pires, 2011).

Following these arguments, initiatives have, for instance, been launched in Sweden and the Netherlands to examine possibilities for relaxing some of these accountability rules and

requirements, and granting civil servants greater discretionary powers. The Swedish Delegation for Trust-Based Public Management — *Tillitsdelegation* — has been founded with the goal of developing a new public management model which will primarily rest on trust in public sector professionals, as apposed to control (Bringselius, 2017). Similarly, in the Netherlands the government has stated its commitment to increasing the possibilities for customization in the provision of public services, and thus, increasing the discretionary powers of civil servants in implementing agencies (ABD TOPConsult, 2019; Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2020).

At the same time, limits on the discretionary decision-making powers of civil servants have been introduced with particular goals in mind too. Civil servants are entrusted with particular powers with the ultimate goal of serving the public. To ensure that those powers are not abused, and the intended goals achieved, requirements to follow specific rules and procedures are often put in place (Bovens, 2007; Pires, 2011; Kwon, 2014). In addition, structuring the work of civil servants around rules and procedures contributes to greater transparency, predictability and consistency in the work of the civil service, which are seen as values in themselves in contemporary democratic societies (Bovens, 2010; O'Donnell, 1998).

There are, thus, arguments for introducing stringent rule-based accountability requirements, as well as for relaxing them. Many of them underpin behavioral expectations regarding the effects of the accountability standard on civil servants. These expectations, however, have not been directly examined rigorously. This study aims to contribute to closing this knowledge gap, by systematically exploring the effect of the specificity of the accountability standard on decision-making behavior in the public sector. Our guiding research question is *how does the specificity of the accountability standard affect decision-making behavior in the public sector?*

To analyze the behavioral consequences of the degree of specificity of the accountability standard, we employ a theoretical model from the field of social psychology. This is, specifically, the most comprehensive socio-psychological account regarding the behavioral effects of accountability, namely *the social contingency model of judgment and choice* (Lerner & Tetlock, 1999; Tetlock, 1992, 1999, 2002). Using this model, we develop expectations regarding the effect of accountability standard specification on decision-making effort and decision impartiality, which are tested using two experiments: an online vignette and a classroom experiment.

In what follows, we first situate and outline the theoretical model on which we base our expectations. We then present the empirical investigation of the research question. After we present our results, we discuss their significance and contribution to the accountability literature. As it will become evident, the manipulation of the specificity of the accountability standard in our study did not result in clear effects on decision-making behavior. The results lend some tentative support to the hypothesis suggesting that accountability for more general accountability standards has positive effects on decision processes in terms of decision-making effort. In contrast, the hypothesis linking more specific accountability standards to outcomes of greater decision impartiality did not find any support. We reflect on what these results signify for the substantive issue at hand, but also for the future of behaviorally informed studies on the effects of public sector accountability.

3.2. Theoretical basis for understanding behavior under accountability pressure

Accountability mechanisms in the public sector have been commonly conceptualized and analyzed with the help of the principal-agent model (Gailmard, 2014; Waterman & Meier, 1998). In this

model, one actor - the agent, undertakes some actions on behalf of another - the principal. Since the model is anchored in rational choice theory, both actors are assumed to be rational utility maximizers and to pursue their own individual, usually conflicting, goals (Waterman & Meier, 1998). In order to keep the agent from pursuing actions which are contrary to the principal's interests, the principal sets up incentives which are aimed at steering the agent's behavior in the desired direction (Gailmard, 2014).

Despite its widespread use, the principal-agent model has proven to be of limited use for predicting bureaucratic behavior (Waterman & Meier, 1998). This is primarily due to two reasons. First, the principal-agent model could be more accurately described as a family of models, and thus a flexible framework for analysis, which requires further specification of assumptions and conditions in order to be a useful tool for hypothesis testing. This flexibility of the model is, therefore, inevitably its greatest limitation, as given the right specification, there is no behavioral pattern the model cannot explain (Gailmard, 2014, p. 92). Second, scholars have noted that the principal-agent model departs in significant ways from the reality of public sector accountability. Specifically, empirical investigations have shown that principals are not always interested in holding their agents to account (Benjamin & Posner, 2018; Schillemans & Busuioc, 2015), and that individual agents are not always inclined to pursue their own selfish interests (Jos & Tompkins, 2004; Koop, 2014).

Scholars have also explored other theoretical avenues for analyzing the mechanisms through which public accountability operates, as alternatives for, or complementing, principal-agent models. Notably, three theoretical approaches have been developed proposing alternative theoretical assumptions. The first one is the stewardship theory, which challenges the principal-agent's assumption of self-interested agents. According to stewardship theory, agents are motivated to act in the best interest of their principals, driven by their commitment to collective values and goals

(Davis et al., 1997; Dicke, 2002). This stream of literature has been primarily interested in uncovering the conditions under which agents act like stewards, however, it has produced some mixed results too (Schillemans & Bjurström, 2019).

The second theoretical approach is the reputational perspective of accountability, which seeks to extend the bureaucratic reputation theory (Carpenter & Krause, 2012) to the analysis of accountability mechanisms (Busuioc & Lodge, 2017). Following this theoretical approach, the behavior of both the account-holder and the account-giver is shaped by their reputational considerations. This theoretical perspective analyzing accountability is arguably the youngest, and has yet to be subjected to rigorous empirical testing. The core motivational assumption postulated by the reputational approach, however, finds its basis in much older theory in the field of social psychology – the social contingency model of judgment and choice (Tetlock, 1992, 1999, 2002). This represents the third theoretical approach investigating the operation of accountability mechanisms. The social contingency model outlines a set of testable hypotheses regarding behavior under different types of accountability pressures, which have been largely supported through extensive experimental research (Lerner & Tetlock, 1999). The relevance of this model for understanding the work of civil servants in the public sector has been emphasized by a number of scholars (for example Hall et al., 2017; Han & Perry, 2020a; Schillemans, 2016). However, it has not yet been used, to our knowledge, to empirically investigate the behavioral responses of civil servants to accountability pressures. We thus, put this theoretical model to the test in a public administration context.

3.3. The social contingency model of accountability

According to the social contingency model, humans can be regarded as “intuitive politicians” whose decision-making behavior is shaped by the social environment (Tetlock, 1992, 1999, 2002). The intuitive politician will anticipate the reactions, and more importantly the objections, of key constituencies in its social environment, and adjust its behavior accordingly, guided by the motivation to maintain a positive image of oneself to audiences to whom it feels accountable (Busuioc & Lodge, 2017; Tetlock, 1992, 1999, 2002).

Decision-makers are considered to be motivated by both symbolic reasons, such as the enhancement of social-image and self-image, as well as obtaining more concrete rewards such as power, wealth and avoiding sanctions (Tetlock, 1992, p.338). These motivations are also echoed in public administration conceptions of accountability (see for example Bovens, 2007; Busuioc & Lodge, 2017; Mulgan, 2003a). Following the social contingency model, to achieve their goals, decision-makers faced with accountability pressures implement one of three coping strategies: the acceptability heuristic, preemptive self-criticism, and defensive bolstering, depending on the accountability context (Tetlock, 1992; Tetlock et al., 1989).

When the decision has already been taken and a pressure to justify the decision is applied only after the fact, the decision-makers are expected to search for arguments in its defense. They will focus on finding arguments that support what already has been done, and thus apply the defensive bolstering strategy (De Dreu & van Knippenberg, 2005; Tetlock, 1992; Tetlock et al., 1989). This strategy, however, is of limited relevance for the purposes of the current study and to accountability pressures in the public sector, as civil servants (should) in principle hold the expectation that they

could be held accountable for their work at any time. It is therefore not a subject of investigation in this study.

When the accountability pressure is introduced earlier in the decision-making process, and thus before the decision has been made, the behavior of the decision-makers is expected to be shaped by the information present in their environment regarding the expectations of the accountability forum. Therefore, when the preferences of the accountability forum are known, the decision-makers will make use of this information to quickly reach a defensible position. Since the decision-makers already have a clear idea what is expected of them, they will not invest additional efforts to search for an optimal solution to the decision-making problem. Thus, in this acceptability heuristic strategy, the decision-makers will use the knowledge about the forum's expectations as a cue to facilitate their decision-making process (Tetlock et al., 1989; Weldon & Gargano, 1988).

Information regarding the expectations of the accountability forum might not always be available, which will mean that the decision-makers will not have an easily available cue in with which they could quickly reach a defensible decision. Since their driving motivation is to maintain a positive image of themselves, they will put an effort to find a solution that they can best defend. Thus, following the strategy of preemptive self-criticism, the decision-makers will take in consideration multiple perspectives, contrast and compare them, and maintain a critical attitude towards the information considered (Tetlock, 1983a; Tetlock et al., 1989; Scholten et al., 2007; Simonson & Nye, 1992).

The degree of knowledge of the expectations of the accountability forum can be paralleled with the specificity of the accountability standard in bureaucratic decision-making. Administrative rules can be considered as expectations for civil servant behavior. They provide guidelines for making

a decision, and are thus cues to build decision defensibility. What is expected of the civil servant is (at least partially) made clear by the rules in place. However, not all rules provide unambiguous guidelines for decision-making. Often they are (purposefully made) vague or abstract, and therefore the civil servant must interpret them (Schillemans, 2012, p.430). They might also only give guidelines regarding a part of the decision-making process, and leave the rest to the discretionary judgment of the civil servant, or be completely absent (Hupe & Hill, 2007; Overman, 2020). The work of civil servants with highly professionalized, complex and technical tasks, for example, is less likely to be defined around the application of specific rules and procedures (Romzek & Dubnick, 1987; Schillemans, 2016). Due to the number of contingencies these civil servants are required to consider in their decision-making processes, and the expert knowledge that they are expected to employ to respond to the problems at hand, they are often given greater autonomy in their work (Romzek, 2000).

When the rules are vague or absent, civil servants do not have a clear cue from the decision-making environment as to what decision is expected from them. Thus, building decision defensibility will require more effort, as the civil servants will have to consider and weigh multiple relevant arguments and factors in the decision-making context. This is, indeed, what we would ideally expect from a civil servant with a highly professional task and great autonomy to perform it (Romzek, 2000; Schillemans, 2016). Based on this application of the social contingency model of judgment and choice in the context of decision-making in the public sector, as well as insights from the public administration literature, we develop two hypotheses regarding the effects of precise rules and general rules on the decision-making behavior in the context of the public sector.

3.4. The case for general accountability standards

The proponents of less stringently defined accountability standards emphasize the importance of conferring public sector professionals with trust and professional autonomy to execute their tasks (Hoecht, 2006; Mansbridge, 2014; Romzek, 2000; Romzek & Dubnick, 1987). Relaxing the rules and procedures for which civil servants are held accountable will grant them with more discretionary space to find the most optimal solutions to the problems presented to them (Molander et al., 2012; Schillemans, 2016), and allow them to address every situation on the basis of its unique circumstances (ABDTOPConsult, 2019; Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2020). This implies that civil servants will actively seek solutions to best respond to the given situation, and thus search and analyze relevant information, consider multiple relevant factors, and base their decisions on a careful weighing of their importance (Brodkin, 1997, p.22; Maynard-Moody & Musheno, 2000). As a result, we would expect that their decision-making process is more effortful and complex than the one of the civil servants who are constrained by stringent rules.

This prediction is also supported by the social contingency model of judgment of choice (Tetlock, 1992), however, for somewhat different reasons. Namely, while precisely specified accountability standards in the form of detailed rules and procedures provide a clear(er) sign of what is expected from the decision-maker, more loosely defined standards leave a room for uncertainty (Tetlock et al., 1989). Motivated by the desire to avoid losing face or face sanctions (Bovens, 2007; Busuioc & Lodge, 2017; Mulgan, 2003a), the decision-maker will employ “preemptive self-criticism”, and thus actively engage in a more complex and effortful decision-making process in order to reach a defensible decision (Tetlock, 1992; Tetlock et al., 1989; Scholten et al., 2007; Simonson & Nye, 1992). Indeed, behavioral research finds that when the expectations of the accountability forum

are not known to the decision-maker, as opposed to known, the decision-maker invests more time in the decision-making process (Klimoski, 1972; Lee et al., 1999), collects more information (Turner, 2001) and also exhibits greater cognitive complexity, or in other words, considers and weighs multiple aspects of the problem (Tetlock, 1983a; Tetlock et al., 1989). Following our parallel between the knowledge of the forum's views and specificity of accountability standard, we expect that less precisely defined accountability standards will lead to more effortful and complex decision-making processing.

H1. The more specific the accountability standard, the less effortful the decision-making process.

3.5. The case for specific accountability standards

Holding civil servants accountable for the application of specific rules and procedures provides structure and predictability in their decision-making process. The consistent application of pre-specified rules and procedures in the processing of each case facilitates fair and equal treatment of cases and clients (Molander et al., 2012; Pires, 2011). Therefore, by following specific rules and procedures, similar cases will be treated similarly, regardless of who the actual decision-maker is.

Specific accountability standards also constrain the decision-making space of civil servants, and thus limit the possibilities for their personal influence over the decision outcome. This guards against biases in the decision-making process, both arising from a conscious pursuit of personal interests (Kwon, 2014; Olken, 2007), or from unconscious stereotypes and biases (Aleksavska et al., 2019; Foschi, 1996). Thus, specific accountability standards would be more conducive to

consistent, unbiased, and thus impartial decision-making processes, than general accountability standards.

The social contingency model of judgment and choice and the behavioral literature also support this reasoning (Tetlock, 1992, 1999, 2002). According to the social contingency model, we would expect that civil servants follow the rules and procedures that apply to the decision-making context, when knowing that they will be held accountable for their implementation. The more specific the prescribed rules and procedures, the clearer the expectations regarding the decision-making behavior of the civil servants. This limits the necessity of the civil servants to engage in preemptive self-criticism, as an acceptability heuristic is readily available. As civil servants do not need to search for additional arguments to defend their decisions beyond what is readily available to them, their decision-making process becomes more standardized (Arkes et al., 2010; Hagafors & Brehmer, 1983; Ordóñez et al., 1999; Siegel-Jacobs & Yates, 1996), and thus more similar across decision-makers (Ashton, 1992; Johnson & Kaplan, 1991). This standardization, as well as the needlessness to seek and introduce additional arguments to justify one's decision, limit the ability of civil servants to introduce personal preferences and biases into the decision-making process. We therefore expect that more specific accountability standards would lead to greater decision-making consistency between civil servants, less influence of personal preferences in the decision-making process, and thus, greater decision impartiality.

H2. The more specific the accountability standard, the more impartial the decision.

3.6. Methodological approach

Experimental scenario and design

In order to investigate our outlined hypotheses, we designed a vignette experiment (Jilke & Van Ryzin, 2017). The experiment presented a scenario in which three short project proposals aimed at reducing plastic waste were described. The participants were asked to evaluate the three projects, and thus give advice on which one to fund and implement. Each participant was randomly assigned to one out of three experimental groups, and thus provided with different instructions as to how to evaluate the projects. One group of participants were asked to make a decision on the basis of their professional experience. This group was, thus, provided no rules to guide their decision-making process, and thus presented a control group. The second group of participants was provided criteria for project evaluation which they were asked to follow, however the criteria were general and required interpretation before they could be implemented. This group was, thus, provided with a general accountability standard. Finally, the third group was provided detailed criteria they were asked to follow in the evaluation of the projects. These criteria provided unambiguous guidelines as to how to evaluate the projects. This group of participants was, thus, provided with a specific accountability standard.

We qualitatively pretested the design of the scenario on a number of master students as well as some colleagues conducting experimental research. Their feedback suggested the experimental design was appropriate and the vignette was seen as realistic, relevant, and perceived as intended.

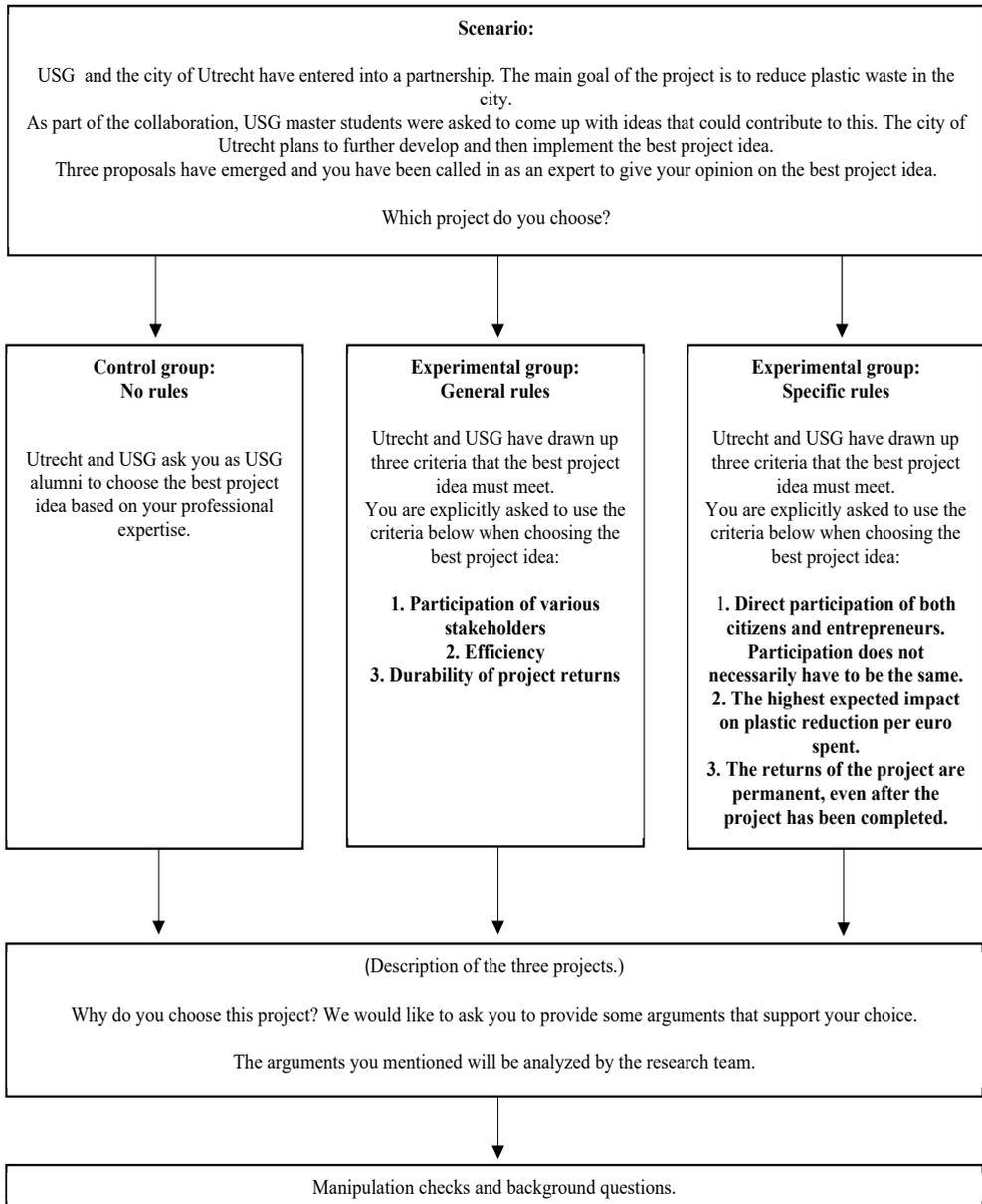
The topic of plastic waste was chosen as a relatively novel and salient one at the time of the experiment, yet not a subject of a political debate and polarization. We expected that the salience of the of topic would have a positive effect on the interest and motivation of our participants to

take part in the study (Boulianne & Basson, 2008). In addition, the novelty of the topic would make it more likely that the information that the participants potentially had regarding plastic pollution was relatively comparable. Finally, the lack of politicization of the topic was seen as beneficial, since it reduced the possibility that particular political and societal groups would have extreme and competing attitudes towards the problem, which could distort our results.

Besides asking the participants to give advice as to which project to implement, we also asked them to provide a justification for their choice. The requirement to justify their decision is an accountability manipulation, which has been well-established in socio-psychological research (Aleksavska et al., 2019; Lerner & Tetlock, 1999). In order to stress the importance of justifying the decision, and thus, to strengthen the accountability manipulation, we emphasized that the justifications will be read and analyzed by researchers working at Utrecht University. The overview of the steps of the experiment is provided in Figure 3.1, while the complete experimental scenario is provided in the appendix.

Measures

The effects on two aspects of decision-making behavior are central in this study, namely, decision-making effort and decision impartiality. The indicators used to measure these two concepts are discussed in what follows.

Figure 3.1 Overview of the experimental design

Decision-making effort. Following H1, the specificity of the accountability standard is expected to affect the effort one invests into the decision-making process. Three behavioral measures are used here to tap into the concept of decision-making effort, namely, decision-making time, justification length, and integrative complexity in thinking. All three measures have been used extensively in previous experimental and behavioral research on accountability as indicators of decision-making effort (see Aleksovska et al., 2019; Lerner & Tetlock, 1999).

The measure of decision-making time captures the total time the decision-maker takes to make the project evaluation requested in the scenario. This includes the time taken to read the instructions and project descriptions, make a choice as to which project to recommend to be funded, and provide a written justification for the choice made. Longer time taken to make a decision signifies greater effort investment (Klimoski, 1972; Lee et al., 1999; Rausch & Brauneis, 2015).

Since the experimental participants provide a justification for their evaluations of the projects, we have a written account of their reasoning behind their evaluations. Decision-makers that provided longer, more detailed, and more thoroughly outlined justifications, arguably put more effort into the decision-making process and in justifying their choice (Koonce et al., 1995; Shankar & Tan, 2006). Thus, by simply looking at the length of their justifications we can evaluate the effort that the decision-makers have put into the process. Justification length is measured in terms of number of characters used.

The third measure of decision-making effort is integrative complexity in thinking, which captures the complexity in the decision-maker's reasoning about the problem (see Suedfield, 2010; Lerner & Tetlock, 1999). Integrative complexity is focused on the structure of one's reasoning, and not its content. It consists of two elements: differentiation and integration. Differentiation refers to the

identification of different dimensions of the problem, while integration refers to the linkages made between them. While higher differentiation and integration both indicate higher integrative complexity in thinking, and subsequently higher effort investment, some level of differentiation is a prerequisite for integration.

The level of integrative complexity that each participant displayed in their reasoning about the problem is determined through an automated analysis of the written decision justifications, on a seven point scale. The analysis is performed with the help of the validated tool developed by Conway III and Conway at the University of Montana (<http://www.autoic.org/>), which uses a pre-defined set of linguistic markers to recognize patterns of differentiation and integration (Conway et al., 2014; Conway et al., 2020; Houck et al., 2014). Before the participant justifications were processed by the automated integrative complexity tool, they were translated from Dutch to English.

Decision impartiality. The proposed benefits of introducing precisely defined accountability standards, as opposed to more general ones, is that they will bring about a greater impartiality in the decision-making process. Impartiality implies similar treatment to similar cases, as well as constraints on the influence of personal preferences of the civil servants in the decision-making process. Thus, two measures are used to capture the concept of decision impartiality, namely, decision consistency and influence of personal preferences.

Decision consistency here refers to similarity in the decision-making primarily *between* civil servants. We are, therefore, interested in the decision consistency *within* each of our experimental groups. The level of consistency in a group of decisions can be determined by investigating their variability: the greater the variability of decisions, the lower their overall consistency and vice-

versa. In numerical variables, the variability is often expressed in terms of variance — or the spread of values around the mean of the group. We, however, do not have a numerical variable, but a nominal one, since the decisions of our interest are recorded as a choice of preferred project. Therefore, to capture the variability of the decisions in each of our experimental groups, we use the concept of “unlikeability” (Kader & Perry, 2007; Perry & Kader, 2005). In simple terms, the coefficient of unlikeability measures how often the observations differ from one another (Kader & Perry, 2007, p. 2). Thus, to calculate the coefficient of unlikeability, each decision in the group is compared to every other decision in that group, and the number of different decisions noted. The coefficient is represented as the proportion of differences from all comparisons made. It thus varies between 0 and 1, with 1 denoting the highest possible unlikeability — all observations are different from each other, while 0 denoting the lowest possible unlikeability — all observations are identical.

Table 3.1 *Overview of measures*

Effect	Measure
Decision-making effort	Decision-making time.
	Decision justification length.
	Integrative complexity of decision justification.
Decision impartiality	Consistency of decisions within an experimental group.
	Overlap of taken decision and personal preference.

To investigate the second measure of decision impartiality, or the influence of civil servants' preferences on the decision, we survey the participants regarding their personal preferences of the three project proposals, after we have asked them to perform an evaluation of the projects. We then compare the expressed personal preference regarding the projects proposals and the project evaluation of each participant, and note when there is an overlap. Table 3.1 provides an overview of all measures.

3.7. Study 1

*Data*¹²

The experiment was run online, on a panel of public administration students alumni from Utrecht University in April and May 2018. In total 633 alumni were invited to participate in the study, out of which, 151 accepted the invitation and completed the study. With this sample, we obtained 75% power to estimate a medium effect size of $f = 0.25$ with $\alpha = 0.05$ (Champely et al., 2020). Therefore, while our sample is reasonably well suited for investigating medium and large effects, it is not powerful enough to capture small effects (Cohen, 1988). More details about the sample composition are provided in appendix.

The experiment was run on a convenience, non-probabilistic sample. It therefore offers limited generalizability. The usage of this sample, however, is adequate for our purposes and offers a number of advantages. First, our study investigates behavioral effects of accountability which are based on psychological mechanisms. It, therefore, aims to detect the presence of general, or

¹² The data from this study is available at: <https://doi.org/10.17026/dans-xka-nure>

universal responses to accountability pressures in public sector context. Second, since the participants have a personal connection with the department of governance at Utrecht University, they can more easily relate to the provided scenario. They are also likely to take the task more seriously for the same reason. Third, all the participants have formal training in public administration, and the majority of them are working as civil servants (information in appendix). This presents a step towards greater external validity of experimental research on accountability, as it presents a more relevant sample of participants for the study of public administration, than the samples of psychology students commonly used in the field of social psychology (Aleksavska et al., 2019).

Manipulation checks

In order to evaluate whether the respondents perceived our accountability standard manipulation as intended, we introduced three manipulation checks. We specifically asked the participants how much influence they felt they had in determining the best project, and whether they thought there were clear criteria for evaluating the projects, to discern the difference between the control group — which did not receive any rules, and the treatment groups — which received rules with different degrees of specificity. Participants in the control group reported experiencing the greatest influence in evaluating the projects, as well as lowest perception of presence of clear rules in the decision-making process, which is in line with our expectations. Differences between the control group and the treatment groups on these two questions are statistically significant. The two treatment groups do not display significant differences on these two questions.

The last manipulation check was intended to capture the difference between the two treatment groups, both receiving rules, but with different degrees of specificity. Participants were asked to

what extent they found the rules to be general or specific. The respondents in the two treatment groups did perceive the specificity of the rules as different, and in line with our intentions. The results from the three manipulation checks are presented in Table 3.2.

Table 3.2 *Manipulation checks results*

	No rules (control group)	General rules	Specific rules
Perception of personal influence on project choice			
Mean	2.18	2.91	2.96
(SD)	(1.14)	(1.38)	(1.22)
N	56	47	48
Summary statistics	F(2; 148) = 6.54, p < 0.01		
Perception of presence of clear rules			
Mean	3.37	2.49	2.35
(SD)	(1.65)	(1.23)	(1.31)
N	56	47	48
Summary statistics	F(2; 148) = 8.02, p < 0.01		
Perception regarding the specificity of rules			
Mean	/	3.83	2.77
(SD)	/	(1.63)	(1.36)
N	/	47	48
Summary statistics	F(1; 93) = 11.84, p < 0.01		

Note. All three questions are measured on a 7-point scale. Perception of personal influence on project choice: 1- very large influence; 7 – very limited influence. Perception of presence of clear rules: 1 – strongly agree; 7 – strongly disagree. Perception regarding the specificity of rules: 1 – very specific; 7 – very general.

The results from the manipulation checks suggest that our participants did perceive our manipulations as intended. They, however, also suggest that the differences between the groups of participants that received the manipulation with general rules and the manipulation with specific rules might be somewhat smaller, or less stark than we originally intended. This could have

implications on the likelihood to observe differences in the decision-making behavior of these two groups.

Results

Decision-making effort. Following H1, we expect that the lower accountability standard specification, the greater the effort investment in the decision-making process. This hypothesis is examined using three measures of decision-making effort: decision-making time, justification length and integrative complexity of thinking. The results of the analysis are presented in Table 3.3 and discussed in turn.

Since the experiment was conducted online, we could not control how or when our respondents decided to complete our experiment. As a result, some respondents could have decided to postpone completion of the experiment, while keeping their browsers open, which could result in untypically large observations of decision-making time which could skew our analysis. These untypical observations are in essence a result of a measurement error, since they do not capture the actual decision-making time of our experimental respondents. They, however, are randomly distributed, and thus do not affect the validity of our findings. In order to minimize the influence of these observations on our analysis, using the method of Cook's distance we identify five outliers, at least four times larger than the mean, and removed them from our sample (Cook, 1977). The results of the analysis of decision-making time displayed in Table 3.3 indicate no significant differences between the three experimental groups.¹³The mean decision-making time per experimental group, however, does follow the expected pattern.

¹³ The values before removing the outliers are $M = 453.99$, $SD = 331.72$ for the group with no rules, $M = 432.98$, $SD = 345.49$ for the group with general rules and $M = 486.47$, $SD = 632.06$ for the group with specific rules. The differences between them are not statistically significant ($F(2; 148) = 0.17$, $p = 0.84$).

Table 3.3 *Effects of accountability standard specification on decision-making effort from Study 1*

	No rules (control group)	General rules	Specific rules
Decision-making time			
Mean	432.21	398.02	347.96
(SD)	(291.58)	(251.59)	(178.50)
N	55	46	45
Summary statistics	F(2;143) = 1.42, p = 0.24		
Justification length			
Mean	373.18	347.76	357.56
(SD)	(350.78)	(254.92)	(314.28)
N	56	47	48
Summary statistics	F (2; 148) = 0.08, p = 0.92		
Integrative complexity			
Mean	2.09	1.82	1.94
(SD)	(1.11)	(0.82)	(0.82)
N	56	47	48
Summary statistics	F(2; 148) = 1.06, p = 0.35		

Note. Time is displayed as seconds spent on the evaluation task while integrative complexity is measured on a scale from 1 to 7.

In terms of justification length we do not observe significant differences between the experimental groups.

Finally, in terms of integrative complexity, the results again do not indicate the presence of significant differences between the three experimental groups.

The results of our analysis do not provide support for H1. We provide reflection on this outcome in the discussion section.

Decision impartiality. To the effect of accountability standard specification on decision impartiality is assessed through the measures of decision consistency and influence of personal preferences. The results of the experimental analysis on the two measures are reported in Table 3.4.

The consistency of the decisions the participants made was assessed by calculating the coefficient of unalikeability for the control and two treatment groups. The coefficient can be interpreted as the proportion of pairs within a group which are unlike. Therefore, larger values of the coefficient indicate lower levels of decision consistency. Contrary to the expectations outlined in H2, the control group, which had the least specific accountability standard, produced decisions with significantly higher levels of consistency than the two treatment groups. In terms of the influence of personal preferences in the decision-making process, the results indicate no significant differences between the three experimental groups.

The results from the decision consistency measure appear to refute H2, while the ones observed regarding the influence of personal preferences display the lack of a significant effect. Taken together, they do not provide support for H2.

Table 3.4 *Effects of accountability standard specification on decision impartiality from Study 1*

	No rules (control group)	General rules	Specific rules
Decision consistency			
Unlikelihood coefficient	0.57	0.65	0.61
N	56	47	48
Summary statistics	$X^2(2, N = 12034) = 27.54, p < 0.01$		
Influence of personal preferences			
Overlap	34	31	25
No overlap	22	16	23
% overlap	60.71%	65.96%	52.08%
Summary statistics	$X^2(2, N = 151) = 1.94, p = 0.38$		

Discussion

The results of the experiment did not provide robust support for our theoretical expectations. This outcome could be due to several reasons. First, the expected relationship could simply not be there, at least not in the form that we expect to find it. There is some tentative evidence for this view, such as the group provided with general rules showing consistently different patterns than expected, for example in the case of the effect of personal preferences and cognitive complexity. Thus, it might be the case that the relationship is not linear as we hypothesized it to be. Second, our instrument could be inadequate for capturing the hypothesized effects. Online experiments are characterized with lower experimental control (Morton & Williams, 2010, p. 533), which could “dilute” the experimental manipulation and lead to non-compliance. After all, the experimental subjects are anonymous and their decision-making within the framework of this online experiment does not bear any direct consequences to them, making the accountability manipulation fairly weak. Third, our experiment might not have sufficient power to capture the hypothesized effects.

The effects that we discuss could be relatively small and thus our sample size might not be sufficiently large for their investigation.

To further investigate the effects of accountability for strict or general rules, and to address some of the potential shortcomings of the first study, we designed a second study on the same scenario yet with a stronger manipulation of accountability and stronger experimental control. Thus, we performed a second round of data collection, with public administration students in a classroom experiment, in which we reduced the experimental groups from three to two to obtain greater statistical power and to focus on the contrast between specific and general accountability standards.

3.8. Study 2

This study employs the same experimental scenario as the first study. Three changes are introduced, however, in order to strengthen experimental control and increase the experimental power. First, the study was conducted in a classroom setting. This was done with the aim to provide greater experimental control and thus achieve greater compliance (Morton & Williams, 2010 p. 532). Second, in order to obtain more experimental power, the experimental groups were reduced from three to two. Therefore, in this experiment participants were placed either in the control condition, where they were provided no rules to follow, or in the treatment group with specific rules, where they were provided with detailed instructions as to how to evaluate the projects. Third, the accountability manipulation was also strengthened in that besides asking the participants to justify their decisions in writing, they were told by the experimenter that some of them would be randomly chosen after they complete the task to explain their decisions to the experimenter, in

front of the other participants. One participant per session was selected randomly to give an explanation of his or her project choice and the reasoning behind the selection.

Data

The experiment was conducted in a classroom setting on a group of students taking an introductory course in public administration in March 2019. 88 students in five different class sessions were offered to take part, out of which four did not give their consent to use their responses in scientific research, and six gave only partial responses. Here we report the results from the 84 students who gave consent to use their responses for scientific research. With this sample, we obtained 59% power to estimate a medium effect size of $f = 0.25$ with $\alpha = 0.05$, or 74% power to estimate a slightly larger effect of $f = 0.3$ (Champely et al., 2020). We therefore, did not obtain a larger power for the statistical test of this sample than the first one. We are, therefore, only able to capture medium and large effects in this study (Cohen, 1988). Nevertheless, due to the changes in the design of the experiment, we obtained greater experimental control and provided a stronger manipulation, both of which reduce the measurement error in this experiment. Additional data regarding the sample is provided in the appendix.

Manipulation check

To investigate whether our manipulation was perceived as intended, we asked our participants whether they thought that there were specific criteria that they had to follow in the evaluation of the projects on a 7-point scale. The two experimental groups did not differ significantly in their responses ($F(1; 77) = 0.30, p = 0.58$). Thus, on the basis of this manipulation check, our participants did not seem to perceive our manipulation as intended. This could be due to several reasons, including misunderstanding the manipulation check question, weak manipulation, or low

statistical power. It should be noted however, that the pattern observed in the data was in line with the our expectations: the group which was provided with specific rules to guide their decision-making process did think there were specific criteria to follow to a greater extent ($M = 2.83$, $SD = 1.58$, $N = 35$) than the group of participants which were not provided any rules ($M = 3.00$, $SD = 1.18$, $N = 44$).

Results

Decision-making effort. We analyze decision making-effort using three measures, as we did in the previous experiment, specifically, with time to make a decision, justification length and integrative complexity. Since this experiment is conducted in a more controlled setting, namely a classroom, we do not observe large outliers when it comes to the time taken to make decision, as we did in the online experiment. However, our measure of the time to make a decision is potentially less precise, since it is self-reported, due to the pen and paper nature of our experiment. The results of the analysis are presented in Table 3.5. As in the previous study, we did not observe significant differences in the time the participants took to complete the task. In terms of justification length we find that the group of participants which had been provided with specific rules provided shorter justifications than the group of participants which were not provided any rules to follow in their decision-making process. This difference was also found to be statistically significant and in line with our expectations outlined in H1.

Regarding cognitive complexity, we do not observe statistically significant differences between the two experimental groups. The results do display, however, a tendency according to our expectations.

Table 3.5 Effects of accountability standard specification on decision-making effort from Study 2

	No rules (control group)	Specific rules
Decision making time		
Mean	7.09	7.72
(SD)	(2.53)	(1.94)
N	42	36
Summary statistics	F(1; 76) = 1.47, p = 0.23	
Justification length		
Mean	417.84	332.72
(SD)	(159.98)	(136.45)
N	45	39
Summary statistics	F(1; 82) = 6.77, p < 0.05	
Integrative complexity		
Mean	2.40	2.33
(SD)	(0.86)	(0.94)
N	45	39
Summary statistics	F (1; 82) = 0.13, p = 0.71	

Note. Time is displayed as self-reported minutes taken to complete the questionnaire while integrative complexity is measured on a scale from 1 to 7.

We thus observe only limited support for H1. We do find support for H1 when it comes to the amount of justification our respondents provided for their decisions, but we do not find support when it comes to decision-making time and integrative complexity.

Decision impartiality. The results of the analysis on the effect of accountability standard specification on decision impartiality are presented in Table 3.6. Similarly, like in the first experiment, the results regarding decision consistency are contrary to our expectations, and the participants in the control group seem to make more consistent decisions than the participants provided with specific rules as an accountability standard. The difference between the two group is also found to be statistically significant. This result provides further evidence for refuting H2.

In terms of the effect of personal preferences on the decision-making process, we do not observe significant differences between the experimental groups. Taken together, the results from the two measures of decision impartiality do not offer support for H2.

Table 3.6 *Effects of accountability standard specification on decision impartiality from Study 2*

	No rules (control group)	Specific rules
<i>Decision consistency</i>		
Unalikeability coefficient	0.55	0.59
N	45	39
Summary statistics	$X^2(1, N = 5430) = 5.25, p = 0.02$	
<i>Influence of personal preferences</i>		
Overlap	27	25
No overlap	17	10
% overlap	61.36%	71.43%
Summary statistics	$X^2(1, N = 79) = 0.49, p = 0.48$	

Discussion

Strengthening the experimental control and experimental power in the second study did not provide much clearer results regarding the behavioral effects of the specification of accountability standards. We discuss the observed outcomes in more detail in the following section and reflect on the results and their methodological and theoretical implications.

3.9. Discussion

Accountability relationships are a ubiquitous presence in the public sector. Even though their existence is justified through the goal of achieving particular performance outcomes, which are closely tied to specific behaviors and values, their ability to do so has not yet been a subject to rigorous investigation (Aleksovska et al., 2019; Bovens, 2010; Dubnick, 2005; Yang, 2012). This study aimed to advance the understanding of the behavioral effects of specifying the accountability standard. A nuanced understanding of the effects of the accountability standards in the public sector will allow for their strategic use to achieve specific goals and to promote particular values (Dubnick, 2005; Schillemans, 2016). It will also provide a better understanding of the consequences of institutional arrangements which inevitably promote a certain type of accountability standard, such as the specific accountability standards fostered by systems with strong checks and balances and judicial review.

The results of our investigation, however, do not provide us the possibility to draw clear-cut conclusions from them. We observed some results that were in line with our expectations, and others displaying patterns contrary to our expectations.

Our results provide some very limited support for hypothesis H1 making the case for general accountability standards. In study 2, we observed that more general standards lead to increased length of decision justifications. We also observed some tendencies in the data in the two studies in terms of decision-making time and integrative complexity, which did not reach statistical significance, but pointed to the expected direction of effect. Therefore, it might be prudent not to discard this hypothesis just yet, and to subject them to further empirical testing.

The picture changes when we look at the hypothesis making the case for specific accountability standards. Here the tendencies mostly go in the opposite direction from our expectations and were significant for decision consistency. Nevertheless, although these results are not strong enough to dismiss our hypothesis altogether, they cast some doubt on the veracity of those expectations. However, as the majority of the statistical tests did not reach conventional levels of statistical significance, we discuss some possible explanations of this outcome.

First, the unclear results could be due to the design of our experimental tool. This experiment was developed after an extensive survey of the behavioral literature on accountability, and modeled on the basis of previous experimental studies (specifically Tetlock, 1983a and Tetlock et al., 1989). Its theoretical and methodological basis are, thus, firmly rooted in the behavioral literature. The experimental tool, nevertheless, has several potential limitations. One of them could be the mode of distribution. Experimental studies on accountability in the domain of social psychology have been primarily conducted in person, often in small groups or in one-to-one experimental settings. Often additional incentives have been attached to them, such as payment for participation or class credits. Our online and classroom experiments could therefore be characterized with lower control and lower stakes than the ones often found in social psychology. Therefore, our accountability manipulation might not have been strong enough, leading to insufficient levels of compliance with the experimental instructions.

Another potential limitation of our experimental design could lie in the operationalization of our core concepts. Specifically, the distinction between our treatment groups with general and specific rules might not have been strong enough, which is exemplified in the relatively close values that the respondents report in the manipulation check of the two groups. Therefore, it is possible that our manipulation was simply not strong enough to capture meaningful effects. In addition, the

interpretation of the provided rules could be partly affected by how the individual respondents perceive them. This, however, might simply reflect the reality of civil servants' work in the public sector.

Finally, our experiments might simply not have been sufficiently powered to capture the hypothesized effects. We saw in several instances tendencies in the results that point to the expected effects, for example for integrative complexity, however, they fell short from reaching conventional levels of statistical significance. Therefore, the effect might be there, however, smaller than we anticipated, and therefore not captured by our experiments. While the experiments reported here have a comparable sample size, and thus power, to the ones reported in the experimental behavioral literature on accountability (for example Tetlock, 1983a; Tetlock et al., 1989 as well as Aleksovskaja et al., 2019 and Lerner & Tetlock, 1999 for overviews), we fail to observe similarly strong effects. This is also a signal that perhaps the effects of accountability are much weaker in public administration contexts, where accountability pressures are a constant presence.

A future study could account for some of the limitations in our experimental design. It could increase the stakes of the experiment and therefore make the accountability manipulation stronger, by for example, organizing the experiment in a one-to-one fashion, or introduce an accountability forum of real-life relevance for the participant (such as one's boss). A future experiment could also include a stronger manipulation of the specification of the accountability standard, which would control for the different perceptions the participants might have of it. The experiment could also be replicated on a larger sample. Observational studies could also be employed to investigate some of the discussed effects, such as the decision-making effort civil servants invest in terms of

time and provided justifications for their decisions in real-life contexts of different accountability standards.

A second possibility as to why our results are less than straightforward, is that the relationships we hypothesize might simply not be there, at least not in the form that we outline here. The social contingency model of judgment and choice on which we base our expectations is not sensitive to iterative relationships (Hall et al., 2017), which are more likely to characterize the work environment of civil servants, and thus, a large part of our respondents. Accountability demands in a public sector setting exert constant pressure (Busuioc & Lodge, 2017), while our study and the theory behind it are based on an acute event. Moreover, the accountability pressures in the public sector have arguably different stakes attached to them than the ones we are able to experimentally manipulate. A failure to properly execute a task in the public sector could lead to various sanctions (Bovens, 2010), such as disciplinary measures and job loss, which are much more impactful than the loss of face that our participants can at worst experience in the context of this experiment. It is therefore possible that the accountability pressures in the public sector operate through different mechanisms than the ones captured by our experiment and theoretical model.

Behavioral approaches present a potentially fruitful avenue for understanding the effects of accountability mechanisms in the public sector (Aleksavska et al., 2019). The existing theoretical models and methodological approaches in the behavioral sciences, however, might not be directly applicable to the study of accountability in the public sector. Further refinement, both theoretical and methodological, is necessary to capture the nature of accountability mechanisms in the public sector. This implies the development of behavioral theories, specific to the public sector, regarding the effects of accountability mechanisms and its characteristics, as well as methodological

approaches which will be able to more adequately capture the way that accountability mechanisms operate in practice.

A behavioral theory of public sector accountability would necessarily need to capture the long-term, iterative nature of accountability relationships in the public sector. Also, such a behavioral theory of public sector accountability would benefit from the incorporation of organization theories explaining how the behaviors of, and in, organizations are affected by the organization's context. Contingency theory for instance postulates that organizational behaviors and effectiveness result from the fit of their characteristics to external contingencies, such as their environment (Donaldson, 2001; see also Olsen, 2013). In this sense, organizational contingency theory and the social contingency model both focus on external contingencies yet differ by focusing on the organization or the individual level. This comparable point of departure could be explored theoretically in order to develop a behavioral theory of public sector accountability. Furthermore, such theory would need to account for the different consequences that civil servants potentially face in their roles as accountees, and consider the stakes they present for them. This will help evaluate the ability of different potential consequences to influence civil servant behavior. Similarly, methodological approaches would need to account for these specifics of accountability mechanisms in the public sector. The long-term, iterative nature of accountability relationships in the public sector could be captured through repeated measurement designs, or through indirect modelling of accountability pressures, by evoking accountability relationships as opposed to directly manipulating them. Indirect modelling of accountability pressures would be also better able to capture and account for the civil servant's perceptions of the stakes and importance associated with particular accountability demands.

3.10. Conclusion

This study aimed to provide a step towards a greater understanding of the behavioral effects of accountability mechanisms in the public sector. We focused on the question “*for what?*” and thus explored the effect of the accountability standard, or what one is being held accountable for, on decision-making behavior. Specifically, we sought to understand how the degree to which the accountability standard is specified affects decision-making behavior in the public sector. Will civil servants’ behavior differ when they are held accountable for the application of rigid rules or when they are held accountable for making a decision based completely on their professional evaluation? Building on socio-psychological theories about the effects of accountability on decision-making behavior, as well as public administration literature, we developed two hypotheses which we tested using two experiments.

One hypothesis made the case that more loosely defined accountability standards might be preferable, since they could potentially stimulate greater effort investment in the decision-making process. In contrast, the other hypothesis supported the case for more stringently defined accountability standards, due to their potential to result in greater decision impartiality.

The results of our investigation offered very limited evidence in support for our hypotheses. Nevertheless, on the basis of the tendencies that our results display, the case for more general accountability standards seems to be stronger than the case for more specific ones. We provide directions for future theoretical and methodological development of the behavioral study of public accountability in public administration.

4. Accountable to whom?

Understanding what makes the
accountability forum matter

4.1.Introduction¹⁴

The work of public sector organizations is constantly scrutinized by numerous stakeholders holding them to account (Busuioc & Lodge, 2017; Koppell, 2005; Willems & van Dooren 2012). This is due to their unique positioning between the stakeholders they serve – the clients and the general public, the central government who authorizes them to operate, and the professional communities and oversight bodies who set and maintain the standards of their work (Bovens, 2007). While this accountability multiplicity is unavoidable in contemporary governance, it creates dilemmas for public sector decision-makers at all levels, from top-level executives to street-level bureaucrats (Lieberherr & Thomann, 2019; Schillemans, 2015). These dilemmas arise from the different, and often conflicting demands the various stakeholders hold, which pushes civil servants to seek ways to reconcile them, or forces them to prioritize some stakeholder demands over others (Busuioc & Lodge, 2017, Schwabenland & Hirst, 2018; Sinclair, 1995; Thomann et al., 2018).

Although the challenge of multiple accountabilities is acknowledged widely in the academic literature (Koppell, 2005; Schillemans & Bovens, 2011), there is very little systematic knowledge about how individual behaviors and decisions in public administration are shaped by the necessity to simultaneously give account to multiple stakeholders (Aleksavska et al., 2019; Busuioc & Lodge, 2017; Yang, 2012). Particularly, the question what factors explain why decision-makers at

¹⁴ Chapter 4 has been published as a research article by Marija Aleksavska, Thomas Schillemans and Stephan Grimmelikhuijsen in *Public Administration Review*:

Aleksavska, M., Schillemans, T. and Grimmelikhuijsen, S. (2021), Management of Multiple Accountabilities through Setting Priorities: Evidence from a Cross-National Conjoint Experiment. *Public Administration Review*. Accepted Author Manuscript. <https://doi.org/10.1111/puar.13357>

Aleksavska designed the study, developed the experimental scenario, collected the data, conducted the data-analysis, presented the data, and wrote the article. Aleksavska developed the UK-panel independently. Schillemans collaborated in developing the scenario together with public sector actors and obtained access to professional respondents in the Netherlands. Grimmelikhuijsen and Schillemans helped to develop the research design and the theoretical positioning, in addition to general supervision and support.

the top or further down the line in the organization prioritize the demands of one stakeholder over others remains unanswered (Busuioc & Lodge, 2017, p. 2; Romzek & Dubnick 1987, p. 235; Schillemans, 2015, p. 438; Schillemans & Bovens, 2011, p. 6). Addressing this research gap is not only relevant from an academic perspective, but also from the perspective of governance and regulation, as understanding the behavior of regulatees when confronted with multiple (accountability) demands, is crucial to effective regulation and enforcement (Ashworth et al., 2002; Boyne et al., 2002).

The literature on public sector accountability suggests two sets of main factors that may drive prioritization decisions. On the one hand, tying in with a conception of accountability as *answerability* (Aberbach & Rockman 1997; Willems & van Dooren 2012), a “traditional” approach would suggest that individual prioritization decisions amidst multiple accountabilities are driven by rational calculations by the individual decision-maker and, in line with principal-agent theory (Gailmard, 2014), the use of hierarchy and sanctions. On the other hand, more modern approaches challenge the dominance of “hierarchical governance” (Hill & Lynn, 2004). In those approaches, in which accountability is seen as the *management of expectations*, other factors such as strategic relationship building and reputation management are said to be ever more important (Busuioc & Lodge, 2017; Romzek & Dubnick, 1987).

Competing logics thus arise from the literature, leaving us in the dark as to what factors are crucial for determining individual prioritization decisions under multiple accountabilities. Against this background, *this study aims to investigate which factors in the accountability setting explain prioritization decisions by civil servants amidst multiple accountability pressures, and what their relative importance is.*

Drawing on the two theoretical perspectives of accountability: answerability versus management of expectations (Acar et al., 2008; O'Loughlin, 1990), we identify four factors which are expected to influence prioritization decisions, and investigate their relative impact. Specifically, we investigate the effects of sanctions, both their likelihood and type, in line with accountability as answerability, and the effects of the forum's expertise regarding the work of the public sector actor, as well as their relationship, in line with accountability as management of expectations.

To be able to make causal inferences about the effect of these factors on priority-setting we designed a conjoint experiment which was fielded on a sample of 600 civil servants in the United Kingdom, and 603 civil servants in the Netherlands. Conjoint experiments allow for the simultaneous manipulation of several variables and measurement of their relative effects (Hainmueller et al., 2014). They are therefore particularly suitable for modeling complex decision-making environments where a number of factors are at play, bringing about greater decision-making realism.

Our analysis indicates that the by far most important factor when deciding which accountability forum to prioritize, with consistent effect, is the threat of sanction. The extent to which the expertise of the forum and the relationship with the forum are taken into consideration by the civil servant is contingent upon contextual factors. This suggests that the classic perspective of accountability as answerability is more relevant than some recent scholarship suggests. We discuss the implications for research but also for practices of governance and regulation.

4.2. The unavoidable multiple accountabilities: Overview of the literature

Public sector organizations are required to give account to a number of external stakeholders, such as their parent departments, clients, professional communities, and even the public at large (Benjamin & Posner, 2018; Romzek & Dubnick, 1987; Thomann et al., 2018;). Each of these stakeholders can influence the work of public sector organizations in their roles as accountability forums. Bovens (2007) defines accountability as a relationship between an accountability forum and an actor, in which the actor has an obligation to inform the forum about his or her conduct, the forum can question the appropriateness of this conduct, and pass judgement which could lead to formal or informal sanctions.

This multiplicity of accountability relationships in the public sector is one of the recurring findings in the academic literature (Hupe & Hill, 2007; Koppell, 2005; Romzek & Dubnick, 1987; Thomann et al., 2018). The multiple accountability pressures stemming from the demands of external stakeholders are experienced at various hierarchical levels within public sector organizations. They are, thus, not only relevant for the work of managers and chief executives (Koppell, 2005; Schillemans, 2015; Schwabenland & Hirst, 2018), but also for street-level bureaucrats (Ewert, 2018; Hupe & Hill, 2007; Lieberherr & Thomann, 2019).

The influence that multiple accountability pressures have on the work of various types of civil servants has often been characterized as problematic (Schillemans & Bovens, 2011). The presence of multiple stakeholders with varying demands creates complexities in the work of the individual decision-maker, since accounting for all of them entails the simultaneous adoption of multiple roles (Ewert, 2018; Shortell et al., 1998). Moreover, the various stakeholder demands can come into conflict, creating dilemmas for the decision-maker as to what is the best course of action

(Thomann et al., 2018). The multiple, conflicting accountability demands have been linked to a number of dysfunctions in the public sector, including cynical blame games, transaction and opportunity costs, organizational paralysis, and diminished efficacy (Koppell, 2005; Romzek & Dubnick, 1987; Schillemans & Bovens, 2011).

These dysfunctions, however, do not appear universally (Schillemans & Bovens, 2011). While making decisions in the face of different and conflicting expectations is indeed a challenge, civil servants have been reported to employ a similar set of strategies to successfully resolve multiple accountability dilemmas. Specifically, scholars have found that when faced with conflicting demands, decision-makers either attempt to develop compromise solutions, or prioritize some demands over others (Busuioc & Lodge, 2017, p. 2; Schwabenland & Hirst, 2018, p. 13; Sinclair, 1995, p. 231; Thomann et al., 2018, p. 315). Alternatively, the less successful ones attempt to appease all accountability forums which may lead to inconsistent patterns of decision-making, and develop a “Multiple Accountabilities Disorder” (Koppell, 2005).

This identified repertoire of strategies for dealing with multiple accountabilities, however, has not been investigated in much detail. Thus, we do not know when decision-makers decide to just implement all accountability demands that come their way and develop a multiple accountability disorder (Koppell, 2005), nor do we know which principles and factors guide prioritization and compromising of demands. The causal links through which the multiple accountability pressures, and their particularities, drive the behavior of public sector decision-makers which ultimately result in some of the suggested outcomes have not been established. This is partly due to the fact that the behavioral responses to multiple accountability pressures have been studied almost exclusively using case studies, with observational or interview data, and this body of literature has

not evolved yet into an integrated and developed research agenda (Aleksavska et al., 2019; Yang, 2012).

As a step towards better understanding of the effects that multiple accountability pressures have on public sector decision-makers, we focus our attention on their prioritization strategy. Understanding which forum's demands get priority can help us get a better grasp of the differences in forums' powers to influence the decisions of public sector actors. In addition, competing theoretical logics may be at play, making it theoretically relevant to investigate their relative impact.

4.3. Competing views on accountability

There are two, partially competing, theoretical perspectives on accountability, which suggest alternative factors which may influence the decision-making behavior of public sector actors: *accountability as answerability* and *accountability as management of expectations*.

Accountability as answerability

The answerability perspective represents the classical, hierarchical view on governance and places great emphasis on the control and oversight aspect of accountability. Accountability is seen as a tool for ensuring that the behavior of public sector decision-makers complies with the expectations set by accountability forums, and that the decision-maker is answerable for fulfilling its mandate (Aberbach & Rockman, 1997; Acar et al., 2008; O'Loughlin, 1990). Accountability pressure is seen as external to the public sector decision-maker and originating from the authority of the accountability forum (Willems & van Dooren, 2012, p. 1014). The relationship between the two

actors is unequal, and the accountability forum holds power over the public sector decision-maker. This view thus presumes clear principal-agent relationship between the two actors (Aberbach & Rockman, 1997; Gailmard, 2014).

The central tool of control, with which the forum can shape the behavior of the public sector actor, is the threat of sanction. However, since a variety of actors act as accountability forums, and draw their authority over the public sector decision-maker from different sources (Willems & van Dooren, 2012), their enforcement style and possibilities for sanctioning could display important differences. Specifically, we argue that the effectiveness of sanctions in shaping behavior would depend on their likelihood and type.

Likelihood of Imposing Sanction. In theory, accountability forums have the possibility to impose a formal or informal sanction to steer and redirect behaviors (Bovens, 2007). In practice, however, scholars observed that forums often fail to use their sanctioning abilities, even in cases of clear misconduct (Benjamin & Posner, 2018; Busuioc, 2009). The reluctance to impose sanctions can be due to the values the forum holds regarding the work of the decision-maker (Benjamin & Posner, 2018; Schillemans & Busuioc, 2015), or due to the possibility of inducing relational costs which would harm the willingness of the decision-maker to comply with the forum's future demands (Braithwaite, 1997; Flinders & Tonkiss, 2016; Greiling & Spraul, 2010) and reduce the overall efficacy of the accountability regime. Thus, the mere ability to impose a sanction might not be seen as a credible threat by the public sector actor if the forum is not likely to implement it. Therefore, we would expect to observe a positive relationship between the likelihood of the forum to impose a sanction for "misbehaving" and the likelihood of the public sector actor to prioritize the demands of a particular forum.

H1: Public sector actors are more likely to prioritize accountability forums who are more likely to punish the actor for non-compliance than ones who are less likely to punish the actor.

Type of Sanction. There is a range of informal and formal sanctions that stakeholders can impose to punish non-compliance, and they are not equally available to all accountability forums. While informal sanctions, such as public shaming, are available to most forums, formal sanctions, such as material consequences, lie primarily in the hands of hierarchical forums. Moreover, not all sanctions bear the same severity, and informal sanctions are generally considered to be weaker than formal ones (Hood et al., 1999). The sanctioning process is often depicted as a process of potential escalation, starting off from lighter informal sanctions which, if not effective, can be substituted by more severe formal sanctions (Braithwaite, 1997). Hood et al. (1999, p. 47) suggest that sanctioning starts with “the ability to shame” and escalates via “lighter weapons” such as certificates and “heavier artillery” towards the “nuclear weapon” of organizational liquidation.

The severity of the sanction the forum uses as a threat is likely to further influence the ability of the forum to ensure compliance through the means of a sanction. When faced with threats of several potential sanctions, the public sector actor is likely to strategically weigh its options and give priority to the forum who threatens to impose the most damaging sanction. Thus, we would expect that formal sanctions will be perceived as more severe and will lead to higher prioritization of forums who threaten to impose them.

H2: Public sector actors are more likely to prioritize accountability forums who threaten to impose a formal sanction than an informal sanction.

Accountability as management of expectations

The modern management of expectations perspective considers accountability to be more than just answerability. It emphasizes the agency of the individual to formulate strategic responses to the different accountability expectations it faces. Here, control mechanisms are seen as only one, and not necessarily even the most important, elements among a larger set of factors that shape the behavior of decision-makers (Acar et al., 2008; Romzek & Dubnick, 1987). The values and interests of the decision-maker take a prominent place in this perspective and act as filtering mechanisms to the various accountability demands (Busuioc & Lodge, 2017). As a result, in the management of expectations perspective, accountability is primarily seen as an internal, rather than external, force (Willems & van Dooren, 2012). It represents self-regulation in the face of external expectations through internal values and strategic considerations.

A key consideration in the formulation of responses to accountability pressures is the building and maintaining of an image of oneself as respectable, trustworthy and professional, vis-à-vis the different accountability forums (Busuioc & Lodge, 2017; Romzek et al., 2012; Tetlock, 1992). This is due to normative reasons, as building a good reputation can be seen as a value in itself, but also strategic ones, since it strengthens ones authority, and consequently powers too. This argument finds its basis in bureaucratic reputation theory (Busuioc & Lodge, 2017), as well as the socio-psychological contingency model of judgement and choice (Tetlock, 1992), both of which postulate largely identical behavioral assumptions, despite differing in their level of analysis.

Building a good reputation, as well as collaborative relations with accountability forums is thus of key importance for the successful management of multiple accountabilities in these theoretical perspectives. However, not all forums are equally valuable for the reputational and strategic goals

of public sector actors. Gaining the approval of the most highly regarded forums would be more beneficial for building esteem, and investing in relations which already have a collaborative nature or potential to do so would be considered more strategic. We thus argue that the forum's expertise and the quality of the relationship with the accountability forum will have an influence on prioritization choices in multiple accountability dilemmas.

Forum Expertise. Professional expertise lends credibility to accountability forums to put forward valid demands regarding the work of the public sector actor, and thus presents a crucial dimension underpinning the efficacy of account-holding (Boyne et al., 2002; Overman et al., 2020). From the perspective of the public sector actor, gaining the approval of an expert forum presents a strategic opportunity, as it can contribute to its professional reputation and esteem (Busuioc & Lodge, 2017; Tetlock, 1992). Indeed, socio-psychologists have found that when the account-holding is performed by actors considered as highly knowledgeable and of high status, accountees tend to invest more effort in their decision-making process (Lerner & Tetlock, 1999; Schillemans, 2016). We would therefore expect that accountability forums with higher expertise will be given priority in multiple accountability dilemmas.

H3: Public sector actors are more likely to prioritize accountability forums who have higher expertise, than ones with lower expertise.

Relationship with the Forum. Many accountability relationships in the public sector have a long-term character. For that reason, building up and maintaining a rapport with the accountability forum is important as it facilitates the continuous management of their expectations (Lieberherr & Thomann, 2019). As Schillemans (2015, p. 438) finds in his interviews with Dutch public sector managers, the active investment in the relationships with the accountability forums is seen as

strategic, since it creates goodwill, it helps to understand the interests of the forum, and to anticipate its reactions. Similarly, Welsh public sector managers have reported that collaborative and supportive relationship with their regulators contributes to the improvement of the achievement of their professional mission (Andrews et al., 2008). Thus, transforming the character of the accountability relationship from that of control and oversight to one of trust and collaboration, eases the communication between the actor and the forum, and gives the actor more freedom and power to pursue its interests (Busuioac & Lodge, 2017; Romzek et al., 2012). Following the strategic considerations of public sector actors, we would expect that they give priority to the demands of the forums with which they have already had positive experience, as that helps to build and maintain positive accountability relationships.

H4: Public sector actors are more likely to prioritize accountability forums with which they have had previous positive experience, than negative previous experience.

4.4. Study design and data

Experimental design

This study was conducted using a conjoint experimental design. Conjoint designs allow for causal testing of the effects of multiple treatment components simultaneously and enable scholars to estimate the relative effect of each of them (Green & Rao, 1971). In this type of experimental study, respondents choose between and/or evaluate a set of alternatives, each having a varied set of attributes (Hainmueller et al., 2014). The method of conjoint analysis is still rare but it is being increasingly used in the fields of political science and public administration in recent years, due to

its many advantages, such as its power to simultaneously investigate several causal claims, as well as due to its simplicity (for example Hollibaugh et al., 2020; Jilke & Tummers, 2018). Conjoint studies limit the effects of social desirability bias, since they allow respondents to justify their decisions using multiple different reasons (Hainmueller & Hopkins, 2015). Furthermore, the choices actors make in the real world, such as choices of public sector actors in multiple accountabilities environments, are often made as a result of trade-offs and interplay of different characteristics of the decision-making environment. Through their design, conjoint experiments account for these trade-offs, creating results with higher degrees of realism (Hainmueller et al., 2014). Finally, the use of conjoint experiments in this context helps diversify the methodological approaches used to study public accountability (Yang, 2012).

In this study, respondents were asked to decide which stakeholder to prioritize from a given set of two profiles (the scenario is provided in appendix). Each stakeholder was described using four attributes, which capture the variables of interest: likelihood of consequences for non-compliance, type of consequence for non-compliance, stakeholder expertise, and previous experience with stakeholder. The overview of the attributes and their levels is provided in Table 4.1. The two stakeholder profiles were randomly generated from the list of available levels of each attribute.

Table 4.1 *Overview of the attributes and their levels*

Attributes	Levels	Operationalization
Likelihood of imposing consequence for non-compliance with demand	High likelihood;	<i>Almost certainly;</i>
	Equal likelihood;	<i>50-50 chance;</i>
	Low likelihood;	<i>Very unlikely;</i>
Possible consequence for non-compliance with demand	Financial;	<i>Financial damage to your organization (fine or budget reduction);</i>
	Reputational;	<i>Bad press, damage to your organization's reputation;</i>
	Relational;	<i>Worsened relationship of your organization with the stakeholder;</i>
Stakeholder expertise	High expertise;	<i>Fully understands the type of work your organization performs;</i>
	Low expertise;	<i>Has very little knowledge about the type of work your organization performs;</i>
Previous experience with stakeholder	Positive;	<i>No struggles at all with this stakeholder in the past, good collaboration;</i>
	Negative;	<i>Many struggles with this stakeholder in the past;</i>

*Data*¹⁵

We performed two rounds of data collection. First, the experiment was run on a sample of 600 civil servants in the United Kingdom, recruited through the online subject pool Prolific.ac (Palan & Schitter, 2018) in July 2019¹⁶. Each participant was asked to make four prioritization decisions, and we thus obtained 2345¹⁷ observations of prioritization choices. A second sample was collected in October 2019 on 603 civil servants in the Netherlands, working in four executive agencies

¹⁵ The data from this study is available at <https://doi.org/10.17026/dans-xt3-k5qc>

¹⁶ The respondents were compensated 0.60 pounds (0.75 US dollars).

¹⁷ Observations where by chance the respondents were asked to choose from two identical stakeholders were removed from the sample.

providing direct financial benefits or grants to citizens. These organizations provide income support (UWV), tax allowances (Belastingdienst-Toeslagen), student loans (DUO), and health care benefits (CAK). The selection and recruitment of participants was done within the organizations themselves. All organizations were instructed to recruit participants who enjoy a certain degree of autonomy in their work, which makes them a subject to accountability pressures. Here again, respondents were asked to make four prioritization decisions, and we obtained 2292 observations of prioritization choices. Table 4.2 provides an overview of the characteristics of the two samples.

Because both samples are different in the way they are collected and composed, it will be hard to make firm inferences from *different* findings in both samples. However, our goal is not to probe any specific differences in each sample, but to assess the *generalizability* of the findings. When the initial findings from the UK sample hold in a second sample taken from a different country, with different mode of data collection resulting in a different composition, this presents a strong case for the generalizability of the findings to other countries and contexts.

Country comparison. From a bird's eyes perspective, the United Kingdom and the Netherlands share several salient characteristics. They are both western developed constitutional monarchies, with politically independent and professional bureaucracies. The Quality of Government indices on bureaucratic professionalism and closedness place the countries close to each other (Dahlström et al., 2015). They nonetheless have at least two differences which are relevant in relation to our research question.

Table 4.2 *Overview of subject characteristics*

	UK sample		Dutch sample	
	N	%	N	%
<i>Manager</i>				
Yes	214	35.7%	31	5.1%
No	385	64.2%	545	90.4%
<i>Supervising employees</i>				
Yes	282	47.0%	113	18.7%
No	318	53.0%	461	76.5%
<i>Client interaction</i>				
Very often	281	46.8%	285	47.3%
Often	117	19.5%	149	24.7%
Sometimes	101	16.8%	71	11.8%
Rarely	53	8.8%	30	5.0%
Never	48	8.0%	41	6.8%
<i>Gender</i>				
Male	132	22.0%	225	37.3%
Female	466	78.0%	338	56.1%
<i>Education</i>				
Secondary school or lower	30	5.0%	29	4.8%
Post-secondary school qualifications	127	21.2%	173	28.7%
Bachelor's degree or higher	443	73.8	372	61.7%
	Mean	Min /	Mean	Min /
	(SD)	Max	(SD)	Max
<i>Age</i>	41.5	20 / 71	47.5	20 / 67
	(10.2)		(10.7)	
<i>Work tenure</i>	9.5	0 / 41	16.3	0 / 46
	(8.0)		(11.1)	
Total N	600		603	

First of all, bureaucracy in the UK has “traditionally developed under the spell of the political domain” (Kuhlmann & Wollmann, 2019, p. 22). Thus, the civil service in the UK is more tightly

controlled and centralized than in other European countries (Hill & Varone, 2014, p. 87). In contrast, the Dutch public sector is traditionally fragmented (Andeweg & Irwin, 2014) and the majority of civil servants, including those surveyed by us, work in agencies at arms' length of central government (Van Thiel & Yesilkagit, 2011). This distinction is relevant for their experiences with multiple accountabilities. Specifically, the accountability cross-pressures, stemming from various accountability forums, our independent variable, would be experienced more strongly by civil servants in fragmented systems, such as in the Netherlands, than by civil servants in more concentrated and politically controlled systems, as in the UK.

Secondly, the UK has a common law tradition enabling policy actors more pragmatic discretion, while the Netherlands has a *Rechtsstaat* system demanding more principled decision-making by civil servants (Bach et al., 2017). This distinction is relevant for the dependent variable in our research, prioritizing between different claims. How much discretion does the civil servant perceive (s)he can take? This relates to impartiality of decision-making, which is found to be considerably higher in the Netherlands than in the UK bureaucracy (Dahlström et al., 2015).

External and ecological validity considerations

Our scenario presents an abstract accountability forum prioritization task, in which the stakeholders are only described with the factors of our interest, while all other context is removed. This approach offers a number of advantages. First, our general scenario allows to investigate various types of civil servants in both countries simultaneously. While multiple accountability dilemmas are likely to present themselves somewhat differently for public sector managers and for street level bureaucrats, our scenario allows the civil servants to associate it with the dilemmas they specifically face in their own workplace. Secondly, from a methodological point of view,

providing more context is likely to introduce noise into the experiment. For instance, a specific identity of a forum is likely to elicit experiences and attitudes towards those specific stakeholders, which we would not be able to account for in our analysis.

We acknowledge that this stylized approach to accountability introduces the possibility that experimental realism is low (Morton & Williams, 2010, p. 265). In order to guard experimental realism, we developed the scenario in consultation with experienced public sector professionals. We also asked participants whether they recognized the type of prioritization dilemma in their work and we found that over 90% of respondents in both samples did (detailed results available in appendix). This suggests that even with a stylized experiment such as ours, experimental realism can be experienced as high.

Analysis

Using the package “cregg” in R (Leeper, 2018), we estimate the marginal mean (MM) for each attribute level in our study. The marginal mean reflects the probability that a certain stakeholder is prioritized given the presence of a particular attribute level (Leeper et al., 2020). It displays the effect of an attribute level with reference to a null effect, or the probability that the stakeholder profile with the particular attribute is chosen completely randomly.

4.5. Results

We present our results in two steps. We first present the main results based on the pooled sample, after which we examine the cross-country variation. Additional analyses examining organizational

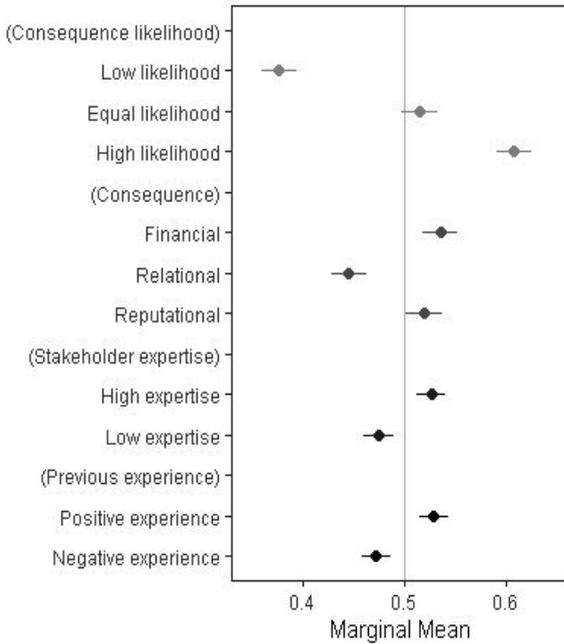
patterns in the Dutch sample and the influence on personal and professional characteristics on prioritization choices are provided in an appendix.

Overall results

The results of the conjoint analysis on the pooled sample are presented in Figure 4.1, where the y-axis presents the attribute levels, while the x-axis presents the marginal means¹⁸. The vertical line at marginal mean 0.5 presents equal likelihood of being prioritized given the specific attribute level. Thus, estimates to the right of the vertical line present higher than equal likelihoods, while estimates to the left of the vertical line present lower than equal likelihoods that the stakeholder with the given attribute level will be prioritized. The closer the marginal mean estimate is to the vertical line, the smaller the effect of the attribute level on the prioritization choice. The horizontal lines running through the marginal means estimates for each attribute level, represent the 95% confidence intervals.

The results displayed in Figure 4.1 indicate that the likelihood of sanction is by far the strongest driver of prioritization choices. Specifically, a high likelihood that the stakeholder will impose a consequence on the public sector actor who does not comply with its demand significantly increases the likelihood of prioritization (MM=0.61, SE=0.01), while a low likelihood significantly decreases it (MM=0.38, SE=0.01). This is in line with our expectations outlined in H1.

¹⁸ Regression models with clustered standard errors per respondent are reported in appendix.

Figure 4.1 *Accountability forum prioritization – pooled sample (N = 4637)*

Note. Figure displays estimated marginal means with confidence intervals set at 95%.

Additionally, we also find that the other elements have some impact on prioritization choices, yet to a much lesser degree. The type of sanction the stakeholder threatens to impose in the case of non-compliance with demands also plays a role. Our analysis indicates that stakeholders who threaten to impose financial (MM=0.53, SE=0.01) or reputational sanctions (MM=0.52, SE=0.01) are significantly more likely to be prioritized than stakeholders who threaten to impose relational sanctions (MM=0.45, SE=0.01). The results partially support H2, in that while we find support that formal financial sanctions will be perceived as more important than informal ones, we do not

find a statistically significant difference between the effect of the financial and reputational sanctions.

Stakeholder expertise also plays a role in prioritization choices. In line with H3, we do find that stakeholders with high expertise are more likely to be prioritized (MM=0.53, SE=0.01) over others. Finally, previous experiences of dealing with the stakeholder do influence the prioritization choices of civil servants in that positive previous experience is given priority (MM=0.53, SE=0.01). This finding is in line with our expectations in H4 (a table with MM estimates is provided in the appendix).

The results of this main analysis underline our theoretical predictions yet also suggest that some of the factors studied are more important than others. It is perhaps not unexpected, yet nevertheless important to see that the threat of consequences is consistently the by far most important factor explaining forum prioritization. The other attributes are also important, yet to a much lesser degree, and this also varies somewhat in the subsamples, which we explore below.

Country comparison

To assess the robustness of the findings in the pooled sample analysis, we evaluate whether they replicate in both country samples independently. Because both samples differ in their composition, specifically, the UK sample contains more managers and highly educated respondents than the Dutch one (see table 2), we performed propensity score matching and extracted subsamples to ensure samples are comparable in terms of subject characteristics (Rosenbaum & Rubin, 1985). The propensity score matching was conducted using the package “MatchIt” in R (Ho et al., 2011), with the method nearest neighbor, discarding unmatched observations from both samples, using a caliper of 0.2. Caliper is a measure of how “different” subsamples are allowed to be and a caliper

of 0.2 has been found to be optimal (Austin, 2011). The matching procedure produced subsamples consisting of 312 respondents from each country which are not statistically different on any of the personal and professional characteristics¹⁹. The subject characteristics post-matching are provided in the appendix. As a robustness check we performed the analysis with both matched and full samples and this yields no substantively different results²⁰.

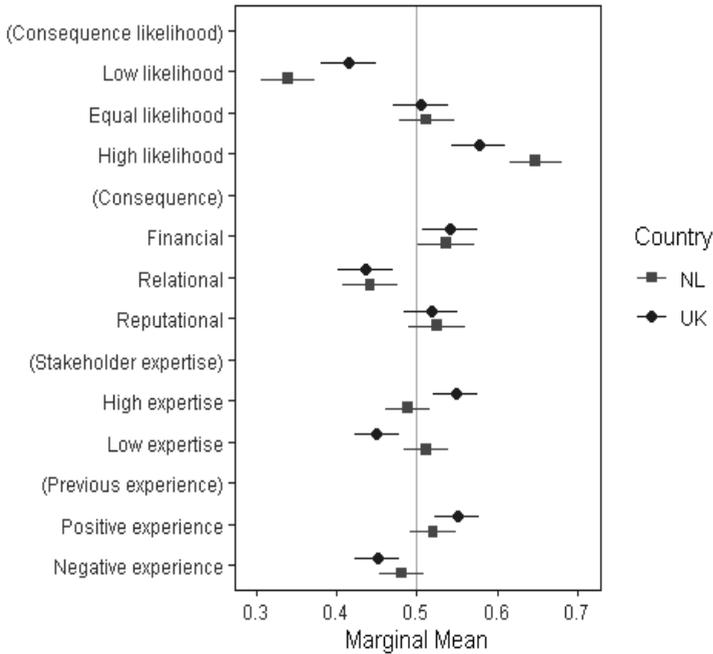
The results of the analysis of the matched samples are presented in Figure 4.2. An omnibus F-test, comparing two nested models, one estimating only the effect of the attributes, and one including a country interaction (Leeper et al., 2020), provides evidence of similarities but also some smaller, yet significant differences in the patterns of prioritization choices of civil servants in the two countries ($F(7; 4866)=6.52, p<0.01$). As figure 2 shows, the prioritization choices of civil servants in both countries are influenced by the threat of sanction in the same way: both groups are strongly driven by the likelihood of sanction. However, this effect is significantly stronger for Dutch civil servants (high likelihood of sanction $MM=0.65, SE=0.02$) than for UK civil servants (high likelihood of sanction $MM=0.58, SE=0.02$). Furthermore, when it comes to the effects of stakeholder expertise and previous experience with the stakeholder, we observe some differences in the responses of the civil servants in the two countries. Specifically, while UK civil servants do prioritize stakeholders with higher expertise ($MM=0.55, SE=0.01$), this is not the case for Dutch civil servants ($MM=0.49, SE=0.01$). UK civil servants also place importance on the previous experience with the stakeholder, and thus give priority to stakeholders with whom they have had

¹⁹ The matching was performed based on all subject characteristics and the extent to which the respondents recognized the dilemma provided in the scenario.

²⁰ The effect of financial sanction is an exception: it is greater in the UK sample than in the Dutch sample, but post-matching this difference disappears.

positive previous experience ($MM=0.55$, $SE=0.01$), although the same tendency can be observed among the Dutch civil servants, the effect is much weaker ($MM=0.52$, $SE=0.01$ respectively).

Figure 4.2 *Accountability forum prioritization per country - matched sample (N = 2441)*



Note. Figure displays estimated marginal means with confidence intervals set at 95%.

4.6. Discussion

Our findings provide support for all four hypotheses, although not to the same degree. The credible threat of sanction increases the likelihood that a forum will receive priority (H1). Furthermore, formal financial sanctions are generally prioritized over informal ones (H2). H3 was also supported

since forums with high expertise are favored over those with low expertise. Finally, positive prior experiences with a forum increases the likelihood of prioritizing that forum (H4).

While all hypotheses are supported by our results, we do find that the strength and the consistency of the effects vary considerably. The likelihood of sanction is by far the most important driver of prioritization decisions in multiple accountability dilemmas, and its effect is robust across subsample analyses. The type of sanction also displays an important and consistent effect: public sector actors see formal financial and reputational sanctions as more important than a damage to the relationship with one stakeholder. These findings are in line with our expectations with regard to sanction credibility, as well as the hierarchy of sanctions (Hood et al., 1999; Schillemans & Busuioc, 2015). Surprisingly, we found no statistically significant difference between the effects of the financial and reputational sanction. One possible explanation of this finding is that since formal sanctions are so sparsely used in practice (Schillemans & Busuioc, 2015), informal ones, such as reputational threats, have gained more gravity. On a more general theoretical note, this suggests that, at least at the level of the individual, the theory of *accountability as answerability* most properly describes how civil servants prioritize their decisions.

In contrast to the consistency of the effects of sanctions on stakeholder prioritization, we found that the effects of stakeholder expertise and the relationship with the stakeholder vary somewhat across contexts. Thus, while stakeholder expertise led to significant stakeholder prioritization in the UK, this effect was not observed in the Netherlands. Furthermore, while in both countries there was a tendency to prioritize stakeholders with whom there was positive previous experience, this effect is much weaker in the Netherlands. Different institutional arrangements as well as values could be the drivers of these differences. The greater fragmentation and stronger commitment to equal treatment of the Dutch civil service, as compared to the UK, could be a part of the

explanation (Dahlström et al., 2015). These tentative explanations, however, should be subjected to further investigation. They suggest, overall, that accountability as *management of expectations* is relevant yet less universal and more context-dependent than accountability as *answerability*.

Our results have at least three implications for the understanding of the behavioral effects of multiple accountability pressures with direct relevance for the study of regulation. First, we observe that there is a clear hierarchy of considerations in multiple accountability dilemmas, and that, like in a Maslow's pyramid of needs, security, and avoiding sanctions, comes first, while building esteem and strategically acquiring powers come after. These findings thus challenge the view of hierarchy as an outdated (Hill & Lynn, 2004; Van Thiel & Yesilkagit, 2011) and ineffective mode of governance (Andrews et al., 2008; Braithwaite, 1997; Flinders & Tonkiss, 2016), and underline that while the modern approaches (Busuioc & Lodge, 2017) provide additional factors for more fine-grained understanding of accountability dynamics, they nevertheless have secondary importance in account-holding and enforcement.

Second, the prioritization choices of public sector decision-makers speak directly to the relative powers of the forums. Evidently, the ones that are able to pose the most credible threat, especially of financial or reputational nature, will be able to exert significant impact on the work of the public sector actor. This might give the upper hand to forums with more aggressive enforcement styles and greater resources to inflict damage. Furthermore, the prioritization of forums with which the decision-maker has had good working experiences signifies that building a good rapport with the accountee could also give the forum a favored position, although less effectively than with the threat to impose a sanction. This type of collaborative relationship between the forum and the actor could be seen as preferable to one of control, as it is less antagonistic, reduces oversight costs (Romzek et al., 2012), and increases the willingness of the public sector actor to comply with the

forum's demands (Andrews et al., 2008). Yet, collaborative relationships could turn into over-comfortable ones, resulting into impunity or regulatory capture (Braun, 2012).

Third, our findings indicate that context likely plays a role in how individuals in the public sector make decisions in multiple accountability dilemmas. Institutional structures, bureaucratic values, and even organizational experiences (see appendix) could introduce variations in the responses to multiple accountability pressures. This calls for tailor-made accountability and regulation models, which are context-sensitive (Schillemans, 2016).

We would like to highlight three limitations of the approach taken in this study. First, while our investigation is based on the two central theoretical views of accountability, it does not capture all of the complexities in the theoretical work on accountability. We opted for theoretical parsimony and placed the emphasis on causal testing of a few central elements through which accountability pressures are expected to influence the decisions of public sector actors, thereby responding to calls for greater methodological rigor and diversity in the study of public accountability (Aleksavska et al., 2019; Yang, 2012).

Second, our experimental scenario presents a stylized task in which the decision-making context is removed. While this approach eases the application of the research tool to various contexts, and thereby facilitates comparability of groups of respondents, it does not account for the nuances of the decision-making context in which a public sector actor is situated. Specific organizational-level dynamics, for example, are not possible to be explored in much detail. Our results could be, thus, treated as a baseline, to which the decision-making in specific multiple accountability environments can be compared. A case in point is a recent study by Hollibaugh et al. (2020) which

shows that in conditions where the forum demands are seen as potentially harmful or unethical, hierarchy and potential sanctions play an insignificant role.

Third, while we use an original dataset, containing two national samples, our data has some limitations. Specifically, our samples are not representative of an underlying population²¹ and they have been collected using different sampling methods in different periods of the year. The differences in the sampling may translate in dissimilarities in sample composition. We accounted for this using the method of propensity score matching (Ho et al., 2011; Rosenbaum & Rubin, 1985). This matching method, however, makes the samples comparable on the basis of measured participant characteristics and is unable to account for unmeasured characteristics that could affect decision-making behavior. The domain of work, the particular stakeholders the respondents are regularly confronted with, the nature of the stakeholder demands, thus again, the context of the decision, could influence prioritization choices in multiple accountability dilemmas.

Future research could explore the potential differences in decision-making strategies under multiple accountability pressures in different sectors, organizations and domains. For instance, do employees in the educational, medical and social sectors make different decisions in multiple accountability dilemmas? Do organizational characteristics and dynamics play a role, and if yes, in which way? Furthermore, investigating the effects of the decision-making context on prioritization decisions, such as the demand ethicalness, decision riskiness and resource scarcity could be of great value for deeper understanding of the mechanisms that drive decision-making amidst multiple accountability pressures.

²¹ Additionally, we are unable to test for non-response bias.

4.7. Conclusion

This study investigated the way public sector actors assign priority to accountability forums in multiple accountabilities dilemmas. The necessity to simultaneously give account to multiple stakeholders has been noted as a universal, and very often problematic phenomenon in the public sector (Hupe & Hill, 2007; Koppell, 2005; Thomann et al., 2018). The way in which the decision-making behavior of public sector actors is shaped as a result of this phenomenon is, however, considerably less understood (Aleksavska et al., 2019; Yang, 2012). Accountability has often been defined either as “answerability” or as “managing expectations”. Our study now suggests that these two processes are both relevant yet not in the same way. Answerability, most notably pre-empting sanctions (financial or reputational), seems to be a universal process preceding other considerations. Managing expectations plays second fiddle, yet seems to be important in explaining more fine-grained contextual differences.

5. Managing multiple accountabilities: The effects of facing multiple forums and conflicting demands

5.1. Introduction²²

Public sector actors are said to operate within webs of accountability (Scott, 2000). They continuously and simultaneously respond to a multitude of legitimate claims coming from a number of different stakeholders, also referred to as accountability “forums” (Bovens, 2010). This has been frequently discussed in the literature under the theme of *multiple accountabilities*. The presence of multiple accountabilities, however, is usually seen in a rather negative light (see Koppel, 2005; Romzek & Dubnick, 1987; Schillemans & Bovens, 2011). Large-scale tragedies, like the space shuttle Challenger disaster (Romzek & Dubnick, 1987), failures of agencies to act consistently (Koppel, 2005), blame games, opportunity costs and purely symbolic accountability (Schillemans & Bovens, 2011), have all been attributed to it. Yet, from the extant literature it remains unclear through which mechanisms multiple accountabilities create these problems, and, thus, how these problems come about (Yang, 2012).

An important hindrance to the systematic empirical examination of the problem of multiple accountabilities, and consequently to deeper understanding of its effects, is its lack of conceptual clarity. Scholars have often evaded defining the concept in clear terms, and have referred to different, yet related phenomena in its discussion. Thus, when talking about the problem of multiple accountabilities, scholars have referred to the multitude of accountability forums (for

²² Chapter 5 has been published as a research article by Marija Aleksovska and Thomas Schillemans in the journal *Public Administration*:

Aleksovska, M. and Schillemans, T. (2021), Dissecting multiple accountabilities: A problem of multiple forums or of conflicting demands?. *Public Administration*. Accepted Author Manuscript. <https://doi.org/10.1111/padm.12763>

Aleksovska made the design of the study, developed the experimental scenario, collected the data, conducted the data-analysis, presented the data, wrote the article. Schillemans collaborated in developing the scenario together with public sector actors and obtained access to professional respondents in the Netherlands. Schillemans contributed to the theoretical section and broader discussion in the paper in addition to general supervision and support (together with Grimmelikhuijsen).

example Caseley, 2006; Voorn, van Genugten, & van Thiel, 2019), the potential conflict that can arise from the presence of multiple demands (for example Hwang & Han, 2017; Koppell, 2005), or most commonly, a combination of the two (for example Bagley, 2010, Thomann et al., 2018). In order to be able to empirically examine the problem of multiple accountabilities and its consequences, and thus advance its understanding, we propose drawing a conceptual distinction between the necessity to give account to multiple accountability forums, and resolving conflicts in accountability demands. Through the means of a vignette experiment, we then investigate the independent effects of these two dimensions of the problem on the decision-making behavior of civil servants.

Multiple accountability pressures are of particular relevance to public sector decision-makers with some discretion, such as executives (Schillemans, 2015), street-level bureaucrats (Ewert, 2018; Hupe & Hill, 2007; Hwang & Han, 2017) and independent regulatory professionals (Lieberherr & Thomann, 2019; Thomann et al., 2018). Our investigation looks into the decision-making behavior of regulatory professionals. As regulators pass judgements regarding the compliance with prescribed norms, which have immediate and potentially severe consequences for the regulatees, investigating their decision-making behavior is of high relevance. By focusing our attention solely on regulators, we are also able to account for the specificities of both the accountability and the decision-making context in which they operate.

Decision-making behavior here refers to the behavior regulators display when making a decision, such as the time and effort spent on the task, asking for help, or gathering information. We focus on micro-level behaviors in order to provide the basis for understanding of some of the large-scale effects that multiple accountability pressures produce in the public sector (Aleksavska et al., 2019; Han & Perry, 2020a; Overman et al., 2020). Building on theoretical and empirical insights from

social-psychology, in combination with public administration literatures on multiple accountabilities, we investigate a number of coping strategies regulators could employ to deal with the accountability pressure. Specifically, we look into “high-effort” coping strategies, or the increased investment of effort into the decision-making process in terms of time, cognitive effort and information search, as well as “low-effort” coping strategies, or the attempts to reduce the accountability pressure through buck-passing, procrastination, and help-seeking (see Green et al., 2000; Tetlock & Boettger, 1994). The usage of both types of coping strategies could signal potential meso and macro problems in the operating of public sector organizations, such as disproportionally large resources allocated to particular tasks (Klingner et al., 2002; Romzek & Dubnick, 1987), decision stalling (Butterfield et al., 2005, p. 339; Murphy & Skillen, 2015, p. 636; Schillemans, 2015 p. 437) and responsibility avoidance (Gilad & Yogev, 2012; Schillemans & Bovens, 2011). Thus, in this study we set to investigate the following question:

RQ: How does the number of accountability forums and the alignment of their demands affect the decision-making behavior of regulatory professionals?

In what follows, we first present the discussions of the “problem” of multiple accountabilities in the literature, its components of forum multiplicity and conflicting demands, and their potential effects on decision-making behavior. We then describe our experimental research strategy, and present the results of our investigation. Finally, we outline the implications of our study in the discussion section.

5.2. The multiple accountabilities problem in the literature

Public accountability, as a concept, captures the obligation of public sector actors to explain or justify their conduct to a significant other (Bovens, 2007). The other is perceived as significant due to its legitimate claim to demand an account, arising from task delegation, or by being directly or indirectly affected by the conduct of the public sector actor. Public accountability is thus a relationship between a public sector actor and an accountee, defined as an *accountability forum* by Bovens (2007; 2010). Besides demanding an account, accountability forums can pass judgements as well as impose formal and informal sanctions (Bovens, 2010; Mulgan, 2003; Thomann et al., 2018).

A number of external actors can have a legitimate claim to demand an account from a public sector actor. Thus, public sector actors commonly face a multitude of accountability forums, who often exert simultaneous pressures, and put forward different demands. The necessity to manage this multitude of expectations and forums has featured in many scholarly discussions on public accountability, often under the theme of *multiple accountabilities*. It has, however, developed a fairly negative reputation, as multiple accountability pressures have been associated with a number of problems and failures (see Romzek & Dubnick, 1987; Koppel, 2005; Schillemans & Bovens, 2011; Thomann et al., 2018). Two prominent studies support this pessimistic view of multiple accountabilities. The first one is the seminal analysis of the 1986 Challenger tragedy, which Romzek and Dubnick (1987) attribute to the unfortunate triumph of political over expert accountability demands. The second one is the case study of the Internet Corporation for Assigned Names and Numbers, whose apparently erratic behavior is explained as the agency's misguided attempt to satisfy all of its accountability demands at once, which Koppell (2005) ominously named multiple accountabilities disorder (MAD). These two studies signify what many other

theoretical and empirical studies argue: multiple accountability introduces complex demands on individuals in the public sector and is said to lead to numerous problematic outcomes (see Bagley, 2010; Schillemans & Bovens, 2011; Thomann et al., 2018).

This critical view on the multiple accountabilities pressures in the public sector, however, might be overstated. Other accounts suggest that multiple accountability pressures are accepted as a fact of life by public sector professionals, and that they are often handled in a fairly routinized fashion, using well-developed and tested strategic coping mechanisms. Analyses of the work of executive public managers in the Netherlands (Schillemans, 2015) and public caseworkers in the United States (Hwang & Han, 2017) have found that these actors routinely use the strategies of transparency, anticipation, information gathering, consultations, and building rapport, among others, in their management of multiple accountabilities. Thus, while the multiple accountability pressures might bring some frustrations, they are generally not perceived as insurmountable problems. In addition, scholars have argued that the simultaneous operation of multiple accountabilities has positive effects on the governance regime as a whole, as it enhances the possibilities to hold power to account (Caseley, 2006; Willems & Van Dooren, 2012). Thus, the literature paints a fairly disjointed picture as to the effects and consequences of multiple accountabilities.

One of the reasons for these disparate findings could be the lack of clear and common conceptualization of the problem of multiple accountabilities. Multiple accountabilities have been commonly conceptualized as a problem of *multiple accountability forums* (for example Caseley, 2006; Voorn et al., 2019), a problem of *conflicting demands* (for example Hwang & Han, 2017; Koppell, 2005), or a combination of the two (for example Bagley, 2010; Thomann et al., 2018). We argue that disentangling and clearly stating these two dimensions of the concept is necessary

for rigorous analysis and deeper understanding of the problem and its associated consequences. Especially so, since, as we argue, the complexities that arise from the multitude of accountability forums and conflicts in accountability demands follow distinct paths, and present challenges of different degrees to regulatory decision-making.

5.3. Disentangling the elements of the concept of multiple accountabilities and their effects

Socio-psychological research shows that the pressure of accountability leads individual decision-makers to think more carefully about the justifiability of their decisions. This is, according to the well-known *social contingency model of judgement and choice*, due to the decision-makers' motivation to gain approval of the audiences they consider as important (Tetlock, 1999), as a means of avoiding negative consequences, building esteem, and gaining power (Tetlock, 1992). This is seen as a fundamental psychological drive operating beneath specific contextual factors such as hierarchy or legal requirements. While considering the defensibility of a decision is a particularly useful consequence of accountability mechanisms in contexts where careful decision-making is expected—such as in many domains of the public sector, like regulation—it can also make decision-making considerably more difficult (Tetlock & Boettger, 1994). This is particularly the case when accountability pressures multiply (Bagley, 2010; Green et al., 2000). As previously discussed, the complexities arising from the simultaneous operation of multiple accountabilities are commonly seen as due to the increase in the number of accountability forums to whom regulators are accountable, or due to the conflicting nature of the accountability demands

regulators face. We examine the effect of each of these dimensions of the problem of multiple accountabilities in turn.

The effect of many forums

The hierarchical structure of governments at face value suggests clear lines of accountability to one salient accountability forum at a time. However, research on regulation in public administration suggests that regulatory professionals in practice are often confronted with many accountability forums representing numerous political, economic and societal actors (Biela & Papadopoulos, 2014; Busuioc & Lodge, 2017). These different actors often hold different interests, and are guided by divergent logics and motivations (Schillemans, 2015; Thomann et al., 2018). The necessity to manage the expectations of these multiple different actors, requires that the regulator take in consideration their goals, interests, and the logics according to which they operate. By carefully considering where the expectations of the forums are coming from, and incorporating those considerations in the crafting of its response, the regulator will have a better chance of eliciting approval from the forums (Tetlock, 1999). Such approval is crucial for maintaining its positive reputation, and, by extension, legitimacy (Busuioc & Lodge, 2017; Carpenter, 2010).

Taking the perspective of multiple actors simultaneously, however, can be fairly complex. In order to be able to respond to the various accountability forums they are faced with, public sector professionals have reported the necessity to continuously shift roles (Ewert, 2018). They therefore have to adapt the information they provide, as well as the manner in which they provide it, to show understanding and acceptance of the forums' perspective, and formulate an adequate response to it. Thus, we argue that even if the accountability expectations are largely identical, the addition of distinct accountability forums creates complexities for decision-makers. The perspectives and

values that every additional accountability forum brings increases the considerations that need to be factored in when formulating the response to the accountability pressure. Therefore, we expect that the mere increase in the number of accountability forums will have an effect on the decision-making behavior of regulators.

The effect of conflicting demands

Beyond the different values and interests that the different accountability forums embody, they can communicate specific demands and expectations of the regulator. These can be reasonably compatible, due to their similarities or their non-overlapping domain, or can stand in direct conflict. The conflicting nature of such demands has often been associated with failures and difficulties in the public sector (for example Koppell, 2005; Romzek & Dubnick, 1987; Thomann et al., 2018).

But how do the complexities stemming from the clash of demands differ from the ones stemming from the different logics that the accountability forums follow? First, a concrete expectation or demand from an accountability forum presents a clearer benchmark of expected behavior. It is, thus, not only about what information is communicated and in which manner, but also about which concrete actions are taken. Such clarity of expectations imposes stricter constraints on the range of actions available to the regulator (Green et al., 2000). This is particularly the case when the conflicting expectations are communicated publicly, which draws increased attention to the behavior of the regulator. Second, when accountability expectations clash, the regulator has the option to fulfill all of them, and go MAD (Koppell, 2005); fulfill some of them, and evidently prioritize some over the others; attempt to reconcile them, and thus fulfill them only partially; or fulfill none of them, and build a justification for a completely different course of action. All of

these scenarios are less than optimal, and can impose significant costs to the regulator. One part of the costs will consist in the efforts invested to formulate an adequate defensible response (Tetlock, 1999), while another will consist in reputational losses the regulator is likely to suffer (Busuioac & Lodge, 2017).

We consider these complexities as different and more burdensome than the complexities introduced by the need to give account to multiple different accountability forums simultaneously. Thus we expect that both the necessity to give account to more than one forum, and the conflicting nature of the expectations, will create complexities in the decision-making of the regulators. However, the need to manage conflicting expectations will present a greater challenge, and thus lead to greater effects on the regulator's decision-making behavior.

H1. Differences in the alignment of accountability demands (compatible versus conflicting) are likely to bring about greater changes in decision-making behavior than differences in the number of accountability forums (single versus multiple).

5.4. Decision-making strategies to cope with multiple accountability pressures

In order to cope with the increased complexities arising from the multiplicity of accountability forums and their potentially conflicting demands, regulators can employ a number of strategies. One way to deal with the mounting pressure is to dedicate more time and energy into the task: complex situations require careful reasoning and crafting of responses. Another way is to attempt to minimize the pressure and reduce the personal responsibility in the situation. Thus, here we discuss and analyze two types of coping strategies: high-effort and low-effort ones.

These two types of strategies have been derived from socio-psychological experimental research on accountability (Green et al., 2000; Tetlock & Boettger, 1994), and adapted to the decision-making context of regulators. Social-psychology has a long tradition of investigating the effects of accountability on decision-making behavior in individuals (Aleksavska et al., 2019; Hall, Frink, & Buckley, 2017; Lerner & Tetlock, 1999), and yet this research has rarely been put into service of aiding the understanding of the decision-making behavior of public sector professionals.

High-effort strategies

The guiding motivation of individuals, following the social contingency model of judgement and choice, *ceteris paribus*, is to gain the approval of relevant audiences (Tetlock, 1992). Thus, individuals will invest sufficient efforts to reach an acceptable and defensible decision, considering the audiences' expectations (Tetlock, 1999). Complex situations, however, will require more effort investment than less complex ones (Green et al., 2000; Tetlock, 1993). As we argued above, increased numbers of accountability forums and incompatibility in expectations present complexities which regulators face. Their presence may thus prompt them to resort to high-effort coping strategies to adequately address them (Green et al., 2000). Here we discuss three such indicators of high-effort investment: decision-making time, decision-making complexity, and information search.

Decision-making time. Civil servants will try to “do their best” (Hwang & Han, 2017, p. 4) to manage the situation at hand. This implies the investment of scarce resources, such as time, to problems which are characterized with greater complexity. In the behavioral literature on accountability the time one takes to make a decision has been strongly linked to the effort invested in the given task (Aleksavska et al., 2019), and is considered as indicative of deeper and more

complex information processing (DeZoort et al., 2006; Kassin et al., 1991; Siegel-Jacobs & Yates, 1996).

Information search. Behavioral research on accountability shows a consistent relationship between the pressure to justify one's decision and information search in the decision-making process (Huber et al., 2009; Ten Velden et al., 2010). It has been argued that this link is due to the fact that more information improves the chances of building better decision justifiability (Tetlock, 1999). This coping strategy has been identified in public administration research on accountability as well, among public caseworkers facing situations characterized with high complexity (Hwang & Han, 2017, p. 5).

Decision-making complexity. Complex situations require complex reasoning about them. Thus, the different arguments and perspectives brought about by the different forums and their diverging demands need to be considered, contrasted or combined in order to arrive at an adequate response to them. As socio-psychological research shows, this triggers more sophisticated cognitive processes, and thus heightens the complexity displayed in the decision-making process (Green et al., 2000; Schillemans, 2016).

Following this discussion, we expect that the increased complexity introduced by the increasing number of accountability forums and conflicting nature of their demands will be met with greater effort investment by regulators in the decision-making process.

H2a. Decision-makers are more likely to use high-effort coping strategies when faced with conflictual accountability demands than when faced with compatible accountability demands.

H2b. Decision-makers are more likely to use high-effort coping strategies when faced with multiple accountability forums as opposed to a single one.

Low-effort strategies

Dealing with multiple accountability demands simultaneously, particularly when they come from different accountability forums and/or are incompatible, can be very taxing. When expectations clash, finding a solution that will be well-received by all parties is especially challenging, and in many cases impossible. These types of situations, thus, can be experienced as very unpleasant and stressful by decision-makers (Bagley, 2010; Green et al., 2000). Instead of confronting the challenging situation directly, they might cave under the pressure and resort to strategies that could reduce the burden of decision-making and their individual responsibility in the situation (Anderson, 2003; Tetlock, 1999). Here we discuss three such strategies: buck-passing, procrastination and looking for help.

Buck-passing. Shifting the burden of making a decision to someone else is the most straightforward way of avoiding accountability. When it becomes apparent that any decision is likely to leave some accountability forums dissatisfied, the opportunity to pass the buck could seem particularly attractive. The high prospect for negative feedback, or even public backlash, could push the regulator to seek escape from the uncomfortable position, and use the opportunity to transfer the case to someone else.

Socio-psychological research has found that when facing complex situations and cross-pressures from multiple accountability forums, decision-makers' likelihood to attempt to pass the buck to someone else increases (Green et al., 2000; Tetlock & Boettger, 1994). This however, is a very extreme form of decision and responsibility avoidance, and might carry significant esteem costs to a professional, such as a regulator. Thus, buck-passing might be regarded as a last-resort strategy of easing the burden of responsibility, and employed only in extreme situations.

Procrastination. A milder form of decision avoidance, is mere postponing of taking an action, as opposed to avoiding taking one altogether. By procrastinating, the regulator temporarily avoids the accountability pressure and alleviates the stress of the situation. Procrastinating also provides an opportunity to better prepare for the potential backlash of the decision, which is reasonably likely in the presence of multiple and competing expectations.

Like in the case of buck-passing, previous socio-psychological research has found that perceptions of decision difficulty and riskiness, as well as simultaneous pressures of accountability forums with opposing views, are more likely to lead to procrastination (Green et al., 2000; Tetlock & Boettger, 1994). In addition, this tactic for coping with accountability pressures has been identified in the work of both street-level bureaucrats and managers (Butterfield et al., 2005, p. 339; Murphy & Skillen, 2015 p. 636).

Although this decision avoidance strategy could be seen as much lighter than buck-passing, it could still bear significant weight in a professional context. This is particularly the case in time-sensitive situations, where prolonging taking an action could lead to missed opportunities or worsening of the state of affairs (Anderson, 2003).

Looking for help. The last and perhaps the mildest form of responsibility alleviation we look into is asking for help from someone else. By asking for help from a colleague or a superior, the regulator can validate his or hers thought process, obtain additional arguments and perspectives to defend a prospective decision, or just attempt to share the burden of decision-making by involving more people in the process (Gilad & Yogev, 2012; Hwang & Han, 2017). When the decision-maker receives help from others, the responsibility for the decision is no longer centered in a single individual, but diffuses among all individuals involved in the decision-making process. It therefore

creates opportunities for avoiding responsibility by hiding behind the group (Kroon et al., 1991) and blame-shifting (Schillemans & Bovens, 2011).

Behavioral experimental research has found that accountability pressures increase cautiousness (Weigold & Schlenker, 1991), and decrease risk-taking (Huber et al., 2009; Kroon et al., 1991) of decision-makers. It is therefore logical that they would ask for a helping hand before making a decision, especially when they are aware that one or more stakeholders are paying particular attention to their actions. In addition, discussions with colleagues and superiors has been identified as one of the crucial coping strategies of social caseworkers in the United States facing conflicting accountability demands (Hwang & Han, 2017 pp. 4-5).

It should be noted, however, that even though we place looking for help under low-effort strategies, we acknowledge that it can also be treated as a high-effort one. Specifically, when decision-makers ask for advice and engage in a thorough discussion with others in order to determine the best course of action given the situation at hand, they invest both time and effort into the decision. Looking for help is only a low effort strategy when responsibility and effort are actually shifted towards others, relieving the decision-makers from the burden of making a decision. Following this discussion, we expect that the increasing decision-difficulty, due to increasing number of accountability forums and/or incompatibility of their demands, will increase the likelihood that regulators pass the buck, procrastinate or look for help.

H3a. Decision-makers are more likely to use low-effort coping strategies when faced with conflictual accountability demands than when faced with compatible accountability demands.

H3b. Decision-makers are more likely to use low-effort coping strategies when faced with multiple accountability forums as opposed to a single one.

5.5. The accountability of regulators

We investigate the case of decision-making by regulators, as they are both a prevalent and relevant case of public sector actors taking important decisions under the pressures of accountability from many forums. Regulators appear as actors on both sides of the accountability relationship: they are simultaneously both account-holders and account-givers (Dudley & Xie, 2019). In their role as account-holders they monitor and enforce the implementation of prescribed rules and standards. To effectively do so, they are vested with powers and resources which allow them to demand information and impose sanctions onto regulatees (Bovens, 2010). Simultaneously, they take up a role of account-givers, as their activities are performed in the name of public interest. Their work is, thus, monitored and evaluated by a range of external actors, including the central government and its relevant ministries, the regulatees, a number of societal and economic actors who bear the direct or indirect consequences of the work of the regulators, as well as the broader public (Biela & Papadopoulos, 2014).

The range of actions that regulators can take is constrained by the law, as well as by the expectations of their accountability forums. Despite these constraints, regulators enjoy considerable discretionary powers, which allow them to make independent decisions. This is precisely why they can be meaningfully held accountable, as only actors who have the possibility to autonomously make authoritative decisions can be actors in accountability relationships (Lindberg, 2013). It is not expected, nor desirable, for regulators to be directly responsive to the wishes of the various accountability forums, as such responsiveness is in direct contradiction with the much-valued regulators' independence. They, nevertheless, cannot be completely deaf to the demands and expectations of the various accountability forums, since they draw their legitimacy from the wide approval and the perceived value of their function (Carpenter, 2010). Thus, while

the expectations of the accountability forums should not be mindlessly followed, they nevertheless need to be heard and managed (Busuioc & Lodge, 2017; Gilad & Yogev, 2012).

5.6. Experimental design and participants

Experimental design

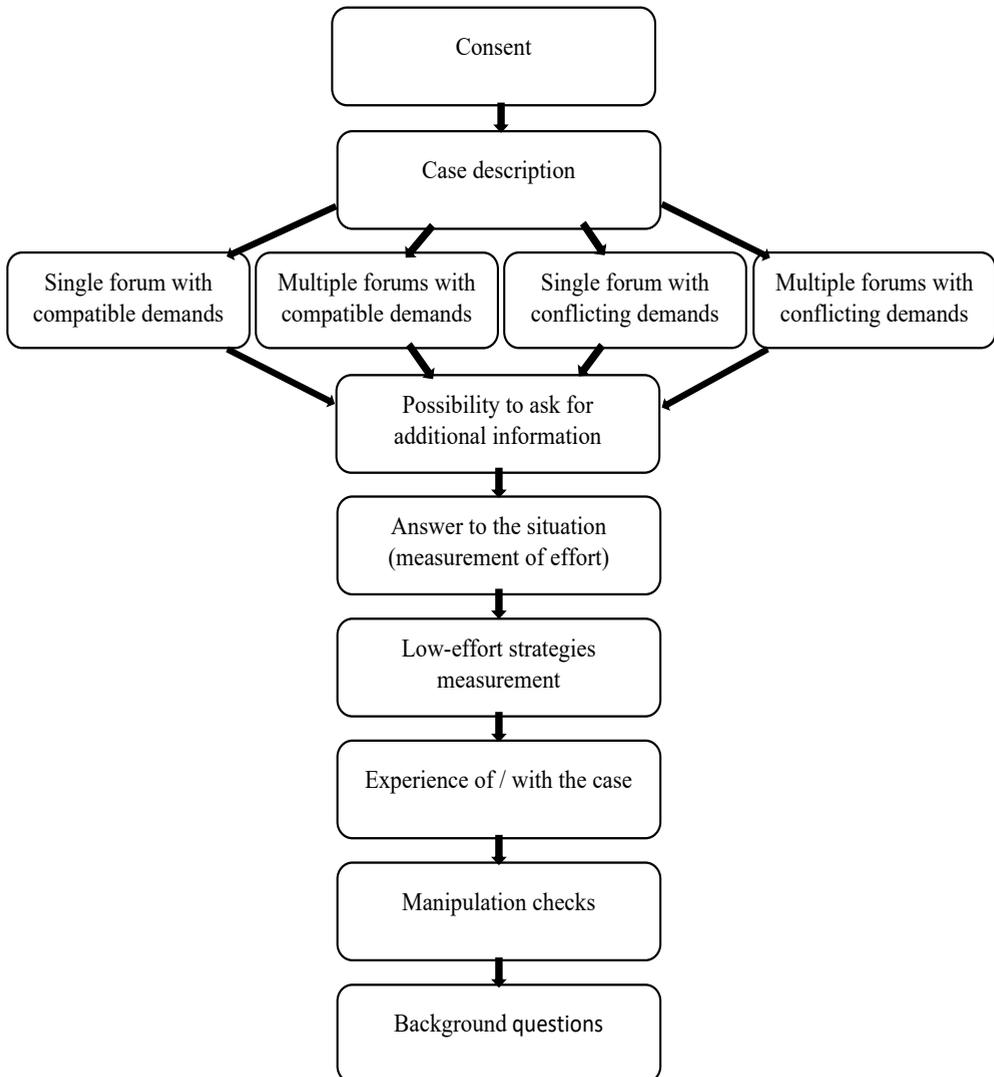
In order to investigate the effect of multiple accountabilities on decision-making behavior, we designed an online vignette experiment. The vignette presented a scenario in which a potentially problematic, but ambiguous situation was presented, to which the respondents were asked to respond in their professional capacity as regulators. The scenario described that the situation received some public traction, and that relevant stakeholders have expressed explicit demands and opinions regarding it. The number of stakeholders (one or two) and the alignment of the demands (compatible or conflictual) were manipulated between participants, making a 2x2 experimental design²³.

After receiving the information about the case and the views of the accountability forum(s), the participants were asked to explain how they would respond to the situation, and were simultaneously provided the option to seek additional information before doing so. They were subsequently asked several questions about their behavioral intentions as well as background information. A flowchart of the experiment is presented in Figure 5.1.

²³ This includes the seemingly counter-intuitive situation where one accountability forum issues conflictual demands. While paradoxical at first sight, this situation has been reported in a number of studies (Pires, 2011, Pollitt, 2003), where it is said to arise from the fact that forums are often complex actors themselves.

The experimental flow had a fixed order as we saw that as the only natural order in which the questions can be asked. We acknowledge that this might give rise to order effects, even though we consider the possibility of the high-effort measures to influence the low-effort ones through priming as relatively low, as they are captured using an open-ended question and direct observation of behavior.

The experimental study was modeled after previous socio-psychological studies investigating the behavioral effects of accountability (Green et al., 2000; Tetlock & Boettger, 1994). These model studies, however, have generally captured situations of social accountability pressures in which students have taken the roles of accountable actors (Aleksavska et al., 2019; Lerner & Tetlock, 1999). Thus, in order to investigate the behavioral responses to multiple accountability pressures of regulators, a significant adaptation of the original studies was required, in order to capture the complexities of the context in which regulators work. To that goal, the experiment was designed in close collaboration with the Dutch Health and Youth Care Inspectorate (IGJ) on whose regulatory work it was modeled. It was further developed together with organizations playing a key role in the professional training for Dutch regulators, the Regulation Academy (Academie voor toezicht), its Knowledge Platform, and the Netherlands School of Public Administration (NSOB). The experimental scenario is provided in the appendix.

Figure 5.1 *Flowchart of the experiment*

Our aim was to achieve a high degree of (mundane) realism in our experiment, so that the participants respond to a realistic situation that they (could) face in their professional environment, and thereby elicit responses from their role as professionals, as opposed to an abstract decision-

making task, which is more likely to prompt universal responses (Bozeman & Scott, 1992). To investigate whether we have achieved this, we asked our respondents the degree to which they recognize the problem of multiple accountability demands outlined in the scenario. All respondents recognized the problem outlined in the scenario, while over 90% of them indicated that they face multiple accountability demands in their work regularly or sometimes. The study's hypotheses were preregistered before any part of the analysis was performed.²⁴

*Measurements*²⁵

Seven indicators are used to capture our dependent variables of interest: three capturing the high-effort coping strategies, and four capturing the low-effort coping strategies.

High-effort strategies

Decision-making time. We adopt the measure of decision-making time to denote the effort our respondents put into responding to the situation provided to them (Kassin et al., 1991; Siegel-Jacobs & Yates, 1996). This is measured as the total time the respondents take to complete the task, from reading the instructions to submitting their written strategy for addressing the situation.

Decision-making complexity. In order to measure decision complexity, we employ the commonly used indicator of integrative complexity. This measure indicates to what extent the respondent incorporates different perspectives in the decision-making process, and attempts to integrate them into one final decision (Tetlock, 1983a). Integrative complexity consists of two dimensions: *differentiation* and *integration*. Differentiation refers to the number of individual characteristics or

²⁴ Preregistration: https://osf.io/926ap?view_only=87c78129b15f4b1a8c919aad810cb499

²⁵ The data for this study is available at: <https://doi.org/10.17026/dans-x22-mcsy>

dimensions of the problem taken into account, while integration refers to the degree to which those are linked (Tetlock, 1983a).

Integrative complexity is measured through manual or automatic coding of the explanations provided by decision-makers to justify their decision. In our study, we ask respondents to explain the decision they took in an open-ended question. Their responses were then translated in English and coded using an automated integrative complexity software: AutoIC²⁶ (Conway III et al., 2014)

Information search. Our respondents were presented with the opportunity to look at additional information before making their decision (Huber et al., 2009; Ten Velden et al., 2010). Here we measure whether they took that opportunity. To see additional information, respondents were advised to tick a box. The additional information provided was a list of measures the inspectorate can implement, and the degree of escalation that each of the measures present. The respondents were informed what the additional information contained, before they decided whether they want to access it.

Low-effort strategies

Looking for help. We ask our respondents whether they would ask for help in the given situation using two survey questions (Hwang & Han, 2017). In the first the source of potential help is a colleague, while in the second it is the boss of the civil servant.

Procrastination. With procrastination, we aim to capture the likelihood that the civil servant postpones or delays taking an action in the given situation (Green et al., 2000; Tetlock & Boettger, 1994). We measure procrastination by directly asking our participants how likely they are to drop

²⁶See: <http://www.autoic.org/>

everything for two full days and help a colleague who asks for urgent help. Due to the urgency of the situation in their own case, deciding to help a colleague before taking an action in their own case would imply tolerating risks to the well-being of a vulnerable group, as well as to the organizational reputation.

Buck-passing. We measure buck-passing by directly asking our participants if they had the opportunity to transfer the case to a colleague, whether they would do so (Green et al., 2000; Tetlock & Boettger, 1994). Details about the measures are provided in appendix.

Participants

The participants of our study are 270 Dutch regulators from over 30 different organizations who responded positively to the invitation to participate in our study in February and March 2020. In collaboration with the Regulation Academy, the NSOB, a closed LinkedIn network for regulators, and the IGJ we invited regulators to take part in the study, using five different survey distribution channels. Our aim was to reach as many regulatory professionals in the Netherlands as possible, working in different organizations and domains. Information about the distribution of the survey and the respondents' characteristics is provided in appendix.

5.7. Analysis and results

Manipulation checks

The experimental manipulations were checked by first asking participants how many stakeholders were discussed in the scenario, and second, whether they perceived the stakeholder demands to be

contradictory or compatible. The participants in the multiple stakeholder groups did note a higher number of stakeholders than the participants in the single stakeholder groups, although not universally ($\chi^2(1, N = 270) = 66.09, p < 0.01$), and the participants in the conflicting demands groups did perceive the stakeholder demands to be more conflictual and less compatible than the participants in the compatible stakeholder demands group ($F(1, 267) = 32.50, p < 0.01$). Thus, our participants perceived our treatments as intended.

High-effort strategies

We first examine the high-effort strategies and thus the evidence for H2a and H2b. Table 5.1 presents the descriptive statistics of the three measures—decision-making time, integrative complexity and information search—and the results of their statistical analysis. In terms of decision-making time, the participants spent approximately seven minutes responding to the problem in the scenario²⁷. The average time the individuals within the different experimental groups took to take their decision does not significantly differ. Similarly, the level of integrative complexity does not significantly differ between the experimental groups either. Unlike for the decision-making time, the data for integrative complexity does however display a trend in line with H2a and H2b: integrative complexity increases with the arising of conflict between the accountability demands and with the increase of the number of accountability forums. Finally, we see relatively high degree of information search in all experimental groups, but no statistically significant differences or meaningful patterns. All in all, we do not find support that conflictual

²⁷ Using Cook's distance, we identified five influential observations (or outliers) of decision-making time, which were four times larger than the mean, and removed them. Prior to removal the mean and standard deviation of decision-making time for the single forum group was $M = 452, SD = 444$, for the multiple forum group was $M = 538, SD = 948$, for the compatible demands group was $M = 496, SD = 810$, and for the conflicting demands group was $M = 506, SD = 728$.

accountability demands (H2a) nor multiple accountability forums (H2b) trigger significantly more high-effort strategies.

Low-effort strategies

Here we examine the low-effort strategies and the evidence our results provide for H3a and H3b. The descriptives of the four measures and their statistical analyses are presented in Table 5.2. The results indicate that facing conflictual demands, as opposed to compatible, and facing multiple accountability forums, as opposed to one, both increase the likelihood of asking for help from a colleague or boss, which is in line with our expectations outlined in H3a and H3b respectively. The likelihood to procrastinate is also affected by the number of accountability forums and the demand alignment. The effect, however, is contrary to our expectations: facing multiple accountability forums, as opposed to one, and facing conflicting expectations, as opposed to compatible ones, reduces the likelihood to procrastinate. The likelihood to buck-pass was generally reported as relatively low in all groups, and not significantly different. Thus, the results provide some support for H3a and H3b, notably when considering the findings regarding asking for help, but also some evidence against the two hypotheses, when considering the findings with regards to procrastination.

Lastly, we examine the evidence in relation to H1. To do so, we look into the effect sizes (η^2) of demand compatibility and the number of accountability forms on our indicators of decision-making behavior. For all of the indicators where the two dimensions of multiple accountabilities were found to have an effect, namely, asking for help from a colleague, asking for help from a boss, and procrastination, the effect of demand compatibility was larger than the effect of the number of accountability forums. Thus, the conflict of accountability demands is more

consequential for the behavior of public sector actors than the multiplicity of accountability forums, which is in line with H1. We reflect on these findings and their implications in the discussion.

Table 5.1 High-effort strategies results from two-way ANOVA and chi-squared analysis

		Mean	SD	N	df	F	p	η^2
<i>Time to make decision</i>								
Number of forums	Single	6m 52s	5m 21s	115	1	0.18	0.67	0.00
	Multiple	6m 57s	5m 33s	150				
Demand compatibility	Compatible	6m 53s	5m 25s	142	1	0.01	0.92	0.00
	Conflicting	6m 57s	5m 31s	123				
Residuals					262			
<i>Integrative complexity</i>								
Number of forums	Single	2.26	1.25	117	1	0.94	0.33	0.00
	Multiple	2.42	1.41	153				
Demand compatibility	Compatible	2.27	1.26	144	1	1.21	0.27	0.00
	Conflicting	2.45	1.43	126				
Residuals					267			
		%		N	df	χ^2	p	
<i>Information search</i>								
Number of forums	Single	70.01%		117	1	0.06	0.81	
	Multiple	67.97%		153				
Demand compatibility	Compatible	69.44%		144	1	0.01	0.94	
	Conflicting	68.25%		126				

Note. Time to make decision is displayed in minutes and seconds, integrative complexity in a 1 to 7 scale, higher numbers denoting higher integrative complexity, and information search in percentage of participants that requested additional information.

Table 5.2 *Low-effort strategies results from two-way ANOVA analysis*

		Mean	SD	N	df	F	p	η^2	
<i>Colleague help</i>									
Number of forums	Single	6.06	1.35	117	1	6.54	0.01	0.02	
	Multiple	6.43	1.06	153					
Demand compatibility	Compatible	6.03	1.46	144	1	13.14	<0.01	0.05	
	Conflicting	6.55	0.73	126					
Residuals					267				
<i>Boss help</i>									
Number of forums	Single	5.38	1.55	117	1	3.50	0.06	0.01	
	Multiple	5.73	1.48	153					
Demand compatibility	Compatible	5.36	1.62	144	1	6.64	0.01	0.02	
	Conflicting	5.83	1.35	126					
Residuals					267				
<i>Procrastination</i>									
Number of forums	Single	3.44	1.51	117	1	3.57	0.06	0.01	
	Multiple	3.09	1.53	153					
Demand compatibility	Compatible	3.49	1.51	144	1	7.90	<0.01	0.03	
	Conflicting	2.97	1.51	126					
Residuals					267				
<i>Buck-passing</i>									
Number of forums	Single	2.38	1.14	117	1	0.33	0.56	0.00	
	Multiple	2.31	1.03	153					
Demand compatibility	Compatible	2.38	1.08	144	1	0.44	0.51	0.00	
	Conflicting	2.29	1.07	126					
Residuals					267				

Note. All measures are on a 7-point scale, with 7 denoting high likelihood and 1 high unlikely.

5.8. Discussion

Our results did not provide support for H2a and H2b, since the effort investment of our participants did not vary as a result of the number of accountability forums, nor the alignment of the accountability demands. We can provide both a methodological and a substantive explanation of this result. This could be partly due to the low experimental control that online experiments

provide: the pressure exerted through the online manipulation of accountability is simply much lower than the pressure felt in real-life situations. The substantive explanation is that the respondents may have acted professionally and just gave their best possible judgments in the experimental task, irrespective of accountability. Considering that as professionals they are frequently faced with complex decisions in multiple accountability settings, this explanation appears highly plausible. A somewhat comparable recent study found that accountability only helped to improve the judgment of junior auditors and had no effects on more experienced auditors (Mala et al., 2018). It may thus be that the behavior of experienced regulators is less susceptible to accountability pressures from those displayed by the experimental participants (commonly students) in socio-psychological studies (Aleksavska et al., 2019).

With regards to low-effort coping strategies (H3a and H3b), we observed mixed results. Regulators did seek more help as the complexity of the situation increased, which was in line with our expectations, however, they did not attempt to pass the buck, and instead of procrastinating, they expediated their actions. These results could be partly driven by social desirability bias allowed by the hypothetical nature of the scenario. However, it is likely that they are indicative of more nuanced strategic responses to multiple accountabilities of regulators, which do not fit precisely in our classification of high and low-effort strategies. We cannot exclude the possibility that asking for help is used here as a low-effort strategy for dispersing responsibility, as originally hypothesized. However, considering the reversed pattern of procrastination we observed in the results, the high-effort interpretation of the asking for help strategy likely applies in this case as well.

Finally, where the number of forums and the alignment of accountability demands did cause a change in the decision-making behavior of our participants, the alignment of accountability

demands had consistently greater impact. This is in line with H1, and unambiguously shows that if multiple accountabilities are in fact a problem, they are most likely a problem of conflicting demands.

All in all, our study makes three specific contributions to the understanding of the multiple accountabilities problem in the public sector. First, we show that disentangling the multitude of accountability forums and their potentially conflicting demands is useful for both theoretical and practical reasons. Of theoretical importance, this study shows that the two factors have distinct effects on decision-making behavior. Practically, the greater size of the effect of demand conflict over forum multiplicity is an important indicator as to where potential issues are most likely to occur.

Second, we show what are some of the coping strategies that regulators are likely to employ in the case of increased complexity due to the increasing number of forums, and the clash of accountability demands. We found that the increasing complexity leads regulators to expedite their actions for resolving the situation, as well as to look for help both among their colleagues and their superiors. These micro behaviors can snowball into a number of more or less desirable meso and macro-outcomes (Overman et al., 2020).

Third, this study demonstrates the potential benefits that psychological research can bring to the study of multiple accountabilities in the public sector: structure in the investigation of behavior, and tools for establishing clear causal links; but also the limitations: inability to precisely predict the decision-making behavior of public sector professionals. This highlights the need for public-administration specific theoretical and empirical behavioral research on accountability, which could account for some of the specific contextual factors which characterize account-giving in the

public sector. Notably, its regularity, stakes, and professional setting, could make public sector professionals more accustomed to dealing with accountability pressures, and thus less affected by them. This dovetails with research on public service motivation (Perry, 1996) and professionalism (Steijn & Noordegraaf, 2014); a link that can be fruitfully investigated in future studies. Specific professional motivations and repertoires could also lead individual regulators to develop different, and perhaps more sophisticated, strategies for coping with multiple accountability pressures than the ones observed in socio-psychological studies.

While this study emanates from discussions in the academic literature on the “problem” of multiple accountability, we believe our findings are also relevant to policy-makers. The study offers at least two practical takeaways for policy-makers. First, the multiple accountability pressures experienced by street-level bureaucrats such as regulators are often portrayed as overburdening (Hupe & Hill 2007; Thomann et al., 2018). Our study suggests that such pressures are real but do not pose a paralyzing threat. This may serve as a source of relief for policy-makers and regulators. These pressures do, however, create additional work for regulators and require the investment of additional efforts and resources. This is exemplified by the help-seeking behavior in accountability situations characterized with higher complexity observed in our study, which implies that to address these situations the effort of multiple regulators is required. This additional effort investment arises solely from the complexities of the accountability environment, and not the task at hand per se. The accountability pressures thus create opportunity costs, as these resources and efforts can be potentially invested elsewhere. When these opportunity costs become so great that they compromise the achievement of core organizational goals, we observe a situation of an “accountability paradox” (Dubnick, 2005 p. 396). Thus, the opportunity costs that accountability pressures create should be closely monitored and evaluated.

Second, our results suggest that prioritization decisions of regulators are in part driven by external accountability pressures. Thus we observe that cases which receive the attention of multiple stakeholders, and those which spark controversy, are treated with more urgency than others. So it is the squeaky wheel that gets the grease, particularly in regulatory contexts where reputation matters (Carpenter, 2010; Gilad & Yogev, 2012). As a result, however, this may increase the chances that problems which do not receive a lot of stakeholder attention, or do not raise controversy may go unnoticed and uncorrected. It is thus important for regulatory agencies to have effective internal processes that prevent seemingly non-salient and non-controversial issues from falling of the tray and receiving too little attention by regulators.

These implications, however, must be read with some care, as our study inevitably has some limitations. First, while we place the strategy asking for help in the group of low-effort ones, as we treat it as a means of dispersing responsibility, we acknowledge that it could be also treated as a high-effort one when seen as an effort to collect views and arguments on how to address the situation the best. Our design does not allow us to distinguish these two interpretations. In addition, we do not explore the relationship between low and high effort strategies, the conditions under which the use of one type of strategies is more likely than the other, nor whether they are mutually exclusive. Second, while the indicators of high-effort coping strategies indicate actual behavior, the low-effort ones capture only self-reported behavioral intentions. Third, even though the respondents recognized the situation in their own work, they were aware that the case is purely hypothetical, and that they would not bear any direct consequences for their decisions. We therefore caution that the results might contain some degree of social desirability bias. Fourth, while the scenario was understood by participants as realistic, it is still abstract and decontextualized. This leaves formal rules and regulations, actual relations with accountability

forums, and decision processes in regulatory agencies out of the equation. And fifth, our experimental design only captures situations where one or two accountability forums put forward specific demands regarding the work of regulatory professionals, while in reality they might face many more. Additionally, our study does not capture nuances of the conflicts between demands, which may vary both in scope, as the issue may be of more or less strategic relevance, as well as in intensity, as the level of conflict between demands can be higher or lower. We should therefore be careful to extrapolate our findings to real world settings. In order to overcome some of the limitations of this study, future research could focus on developing more direct ways of measuring the behavioral responses to multiple accountability pressures, potentially employing observational and document studies, as well as field and natural experiments in which actual decision contexts can be incorporated.

5.9. Conclusion

By means of a realistic vignette experiment, this study investigated the causal effects of multiple accountability pressures on the decision-making behavior of regulatory professionals, conceptualized as high and low-effort coping strategies (Green et al., 2000; Tetlock & Boettger, 1994). We argued for theoretical disentangling of the two dimensions of the multiple accountabilities “problem”—the number of accountability forums, and the alignment of their demands—and empirically examined their independent effects. Our investigation found that the number of accountability forums and the alignment of their demands have no effect on the likelihood to employ high-effort strategies, measured as the time spent on a task, integrative complexity and information search. They did have an effect on the likelihood to employ low-effort

strategies, in terms of asking for help and procrastination, but not on the likelihood to pass the buck, and not entirely in line with our expectations. Finally, where these two dimensions did have an effect on the decision-making behavior of regulators, the effect of demand non-alignment was consistently greater than the effect of forum multiplicity.

6. Conclusion: Understanding accountability through a behavioral lens

6.1. Why we need to understand the behavioral consequences of accountability in the public sector

Accountability mechanisms have long been considered as tools for steering behavior in the public sector. Their operation has often been seen as a means to achieve certain normatively desirable outcomes, such as the proper execution of delegated tasks, the prevention of abuse of power, as well as the improvement of performance (Bovens, 2007; Bovens et al., 2008; Dubnick, 2005). The evidence of the actual effects of accountability mechanisms in the public sector, however, does not fully correspond to these optimistic expectations. In reality, accountability has been associated with a range of effects, many of them neither entirely anticipated, nor seen as desirable (e.g. Benjamin & Posner, 2018; Flinders, 2011; Koppell, 2005; Schillemans & Busuioc, 2015). A case in point is the classic analysis of the 1986 Challenger tragedy (Romzek & Dubnick, 1987), with which this thesis started, which illustrates how the simultaneous and conflicting accountability pressures NASA was facing pulled the agency in different directions. This forced the agency's management to prioritize some accountability demands over others, and in response to the accountability pressure to launch hurriedly, to tolerate technical risks that were simply too high, ultimately leading to the disaster.

This discrepancy between expected and observed effects of accountability is, at least in part, a result of the lack of systematic empirical investigation of the causal effects of accountability mechanisms in the public domain (Flinders, 2014; Yang, 2012, 2014). The anticipated positive effects of accountability in public administration are much more strongly rooted in theoretical and normative discussions than in empirical evidence (Bovens et al., 2008; Dubnick, 2003, 2005). While empirical evidence supporting the claims of normatively desirable effects of accountability certainly exists (e.g. Demir et al., 2019; Olken, 2007), it is by no means comprehensive enough to

be considered conclusive. This has, understandably, invited a lot of research into the failures, deficiencies, and problematic outcomes of the operation of accountability mechanisms (e.g. Bevan & Hood, 2006; Halachmi, 2014; Mulgan, 2014; Thomann et al., 2018), challenging the universal positive connotation of accountability. Many of these studies, however, have focused primarily on single cases and instances of failures (like the Challenger tragedy study), without emerging into a cumulative body of knowledge or a focused research agenda. The blurry conceptual boundaries of accountability and the fairly liberal use of the concept have not helped in this regard (Flinders, 2014; Pollitt & Hupe, 2011). Thus, while there is at present a wealth of theoretical, normative, conceptual and empirical research on accountability, offering many useful insights, there remain important gaps in the understanding of the consequences of accountability (Yang, 2012, 2014).

The aim of this dissertation has been to help fill some of these gaps by focusing on accountability effects on the individual or micro-level in the public sector. More specifically, the dissertation focused on the decision-making behavior of civil servants with some discretionary powers facing accountability. Tracing the individual responses to accountability pressures allows us to locate the foundations of some of the associated meso- and macro-effects of accountability (Overman et al., 2020), such as improved organizational performance, or unsuccessful NASA space shuttle launches. In addition, it makes it possible to benefit from the insights of the rich behavioral, and particularly socio-psychological, research on the subject (Hall et al., 2017; Lerner & Tetlock, 1999; Schillemans, 2016). The overarching research question that has guided this research project is:

How do accountability mechanisms shape decision-making behavior in the public sector?

The central question was explored in four sub-projects, each guided by a more focused research question. The first sub-question, examined in Chapter 2, was: *What insights can be drawn from the behavioral research on accountability for its study in the domain of public administration?*

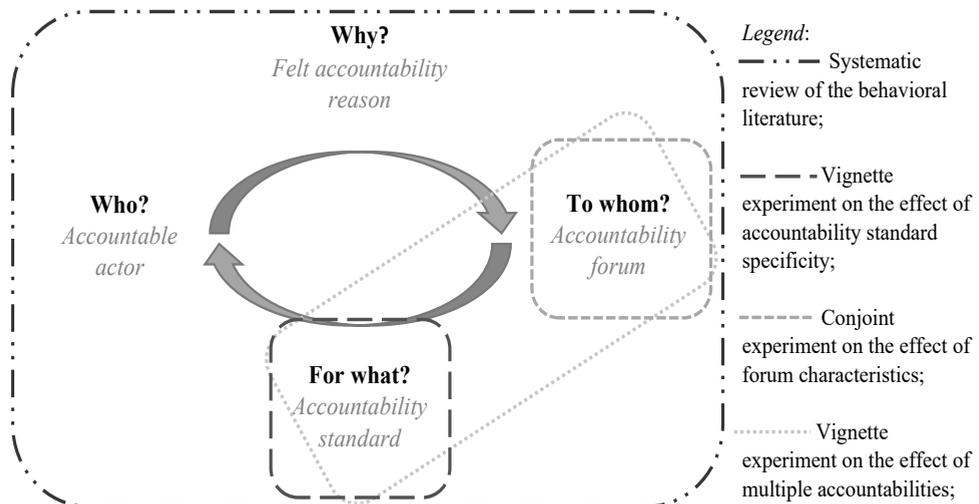
This sub-project thus examined the behavioral literature and its insights and findings with regards to the causal effects of accountability mechanisms in relation to key accountability issues in the public administration literature. On the basis of a systematic literature review, 266 experiments from the behavioral sciences examining the effects of accountability on a range of outcomes were identified. One part of these experimental studies examined the effects of introducing an accountability mechanism, thereby contrasting outcomes where accountability is present and where it is absent. The rest of the experimental studies examined the effects of specific features of accountability mechanisms, and thus contrasted the outcomes of accountability mechanisms with different characteristics in terms of the *timing* of accountability, the accountability *standard*, or the *relationship* between the accountable actor and forum. Overall, it showed that accountability can have many effects on the individual level, many of which are relevant to issues raised in public administration.

The systematic literature review was followed up by three experimental sub-projects that investigated the effects of specific configurations of accountability mechanisms in realistic public administration contexts. They have focused on the accountability standard (for *what* are decision-makers held accountable?) and the accountability forum (*to whom* are decision-makers accountable?) in more detail. The second sub-question, guiding the design of a vignette experiment presented in Chapter 3, thus was: *How does the specificity of the accountability standard affect decision-making behavior in the public sector?* The third sub-question focused on the accountability forum, and was: *Which accountability forums get their demands prioritized?* The experimental sub-project, reported in Chapter 4, examined the priority different accountability forums receive, on the basis of their characteristics and their relationship with the accountable actor, using a conjoint experiment. The fourth and last sub-question, presented in Chapter 5 of the

dissertation, combined the questions of *for what* and *to whom*, and was: *How does the number of accountability forums and the alignment of their demands affect the decision-making behavior in the public sector?* Thus, this sub-project investigated the decision-making behavior of regulators contingent upon the number of accountability forums they are facing, and the compatibility of the demands they put forward, again using a vignette experiment.

Figure 1 below offers a visual presentation of the scope of the four sub-projects presented in the dissertation, depicted as zooming in on specific elements of the accountability mechanism.

Figure 6.1 *The research focus of the dissertation*



In the remainder of this concluding chapter, the results of the analyses of the effects of accountability mechanisms in the public sector conducted as part of this dissertation will be presented and reflected upon. I first present the key findings of the four sub-projects. Then I discuss

the limitations of the conducted empirical research. Finally, I put forward a broader reflection on the behavioral approach taken in this dissertation, and offer some ideas for its further theoretical development.

6.2. Accountability and decision-making behavior: Universal and context-dependent effects

The survey of the behavioral experimental accountability literature presented in Chapter 2 suggests that accountability mechanisms are a powerful tool for influencing decision-making processes, behaviors and outcomes. Not only their introduction, but also the variations in their characteristics (in terms of timing, standards, and the relationships between the actor and the forum), have been found to influence, among other things, the likelihood that decision-makers invest more effort into the decision-making processes, take risks or collaborate, as well as shape their substantive judgments and decisions. These findings from the broader behavioral literature served as baseline expectations for the effects of accountability in the more specific context of the public sector.

What the experimental investigations of the operation of accountability mechanisms in the public sector indicate is that certain behavioral responses to accountability observed in the broader behavioral literature persist in the public sector context, while others do not. For instance, the results from the experimental research confirm the finding from the behavioral sciences that the demands of authoritative and assertive forums are taken up more seriously than others. They however do not provide evidence that the decision-making effort of civil servants varies strongly as a result of changes in the accountability pressures. These two key findings are discussed in more detail in what follows.

Finding 1: The loudest, most threatening accountability forums receive most attention. Not all accountability demands were attributed equal importance by the accountable actors in the different studies. As we saw in Chapter 4 and Chapter 5, some accountability demands were treated with more urgency than others, depending on who issued them and in what way. The results presented in Chapter 4 indicated that the threat of sanction by the forum increased the priority with which the accountability demand was handled. Similarly, in the experiment presented in Chapter 5, the accountable actors decided to expediate their actions in the face of increasing involvement by their accountability forums and the emergence of conflicting expectations. This may not sound surprising at face value, however, many recent studies in public administration have emphasized the effects of informal, social, horizontal or other forms of non-hierarchical accountability (Brummel, 2021; Romzek et al., 2012; Schillemans, 2011). The experiments in this manuscript underscore the persisting impact of traditional, tougher forms of accountability. These findings are in line with the conclusions from the behavior literature, which indicate that the identity (Mero et al., 2007; Tan & Jamal, 2006) and the communication style (Cohen & Trompeter, 1998) of the forum will influence the degree to which the forum's demand is taken seriously by the accountable actor.

Finding 2: Variations in accountability pressures, yet stable decision-making effort. One somewhat puzzling pattern emerging from the results presented in this dissertation, is that the investment of effort in the decision-making process is not particularly sensitive to variations in the accountability pressures in the context of the public sector, at least in these experiments. We observe this in Chapter 3, where the consequences of varying the specificity of the accountability standard are examined, as well as in Chapter 5, where the impact of the number of accountability forums and the alignment of their demands are investigated. In both experiments, we see

remarkable stability in the effort our participants invest in their decision-making processes in terms of decision-making time, complexity of thought, and information search, and only observe positive effects in terms of the provision of more comprehensive justifications. These findings from the specific public sector context contrast with the accumulated evidence from the broader behavioral sciences presented in Chapter 2 that accountability is a powerful tool for influencing individual effort investment. Thus while individuals in mostly decontextualized settings have consistently been found to significantly adapt their effort-investment to the accountability circumstances, and therefore increase their decision-making effort when they face forums whose expectations are unspecified (Turner, 2001; Tetlock, 1983; Tetlock et al., 1989), or who hold heterogeneous or conflicting views (Green et al., 2000; Shankar & Tan, 2006), the same does not appear to hold in more realistic experimental scenarios in public sector settings.

The combination of success and failure in replicating some of the effects of accountability mechanisms from the broader behavioral sciences in the specific public sector context, suggests that general behavioral theories of accountability might not be the silver bullet to understanding the effects of accountability in the public sector. Instead, the application of behavioral science to the study of public sector accountability appears to be more of a mixed blessing: useful in some respects, but insufficiently sensitive to important contextual dynamics. This will be discussed in more detail below, however only after the limitations of the studies are discussed, as the combination of success and failure could also be a product of the (inevitable) limitations of the research that was conducted.

6.3. Limitations of the presented research

The research conducted in this dissertation inevitably has certain limitations which may to some extent explain why certain effects were, or were not, established. Many of these limitations arise from the methodological approach taken to investigate the central research question. First, the ambition to primarily recruit public sector professionals as participants in the experimental studies has come with certain costs in terms of (control of) the sample size. As public sector professionals constitute a much smaller population than ordinary citizens, and a much more exclusive one than students, large samples of participants have been difficult to achieve. In addition, the collaboration with public sector organizations in the recruitment of participants has entailed a certain loss of control over the recruitment process, both in terms of *who* specifically gets to be invited to participate, and *how many* public sector professionals end up participating in the studies. Thus, some of the hypotheses of the individual studies were not confirmed, potentially as a result of modest sample sizes.

Second, even though the studies presented in this dissertation have aimed to create high experimental realism, they still present hypothetical scenarios in which the accountability manipulation is artificially introduced. As these hypothetical scenarios do not carry the same weight as real situations where accountability pressures apply, it is conceivable that the decision-making behavior of public sector professionals observed here departs in some ways from the behavior we would see if the situation occurred in the real world. It is easy to imagine that real and consequential forms of accountability may have more powerful effects on individuals than accountability manipulations in artificial settings.

Third, many behaviors of public sector professionals which are of interest to public administration scholars are difficult to observe directly in an experimental setting. A good example could be

procrastination, which happens daily in many offices and is likely to be an effect of accountability pressures, but is difficult to re-create in an experimental setting within limited time. As a result, this thesis has made use of self-reported behavioral intentions as measures of behavior in certain instances, which can be susceptible to social desirability bias. This may have prevented us from uncovering some of the socially undesirable effects of accountability.

Fourth and last, the decision-making context in which public sector professionals operate is highly complex, and modeling it in an experimental study inevitably implies its simplification. The experimental studies presented here thus focus on only a few relevant factors from this rich decision-making context, and leave a large number of other contextually relevant factors out of the equation. In that sense these studies are, although more realistic for public administration settings than the original socio-psychological experiments, still marked by artificiality.

From a methodological point of view, the exploration of the behavioral effects of accountability mechanisms could benefit from further methodological diversification. This dissertation has taken a step in that direction, by introducing several experimental studies in an area of inquiry dominated by qualitative case studies. However, just like case studies, experiments have their limitations, as the preceding discussion demonstrates. The only way the limitations of any chosen method can be overcome, and the validity of findings strengthened, is by practicing methodological pluralism and triangulation (Yang, 2012).

6.4. Behavioral science and public sector accountability: A mixed blessing

To further advance the understanding of the decision-making behavior of individuals within the public sector, this dissertation has drawn extensively on theoretical, empirical and methodological

insights from the behavioral sciences in combination with public administration literatures. The social contingency model of judgement and choice (Tetlock, 1992, 2002) and its many empirical examinations and extensions (Lerner & Tetlock, 1999), have provided the theoretical basis for much of the work conducted in this dissertation, while the experimental method has been used as a key tool for testing the causal effects of accountability. Here I reflect on their usefulness to aid the study of accountability in the public sector, and thus discuss where they excel and where they fall short.

Major contributions of behavioral science: Basic behavioral model and sophisticated toolbox for causal inference

Overall, this dissertation suggests that the behavioral sciences can make two key contributions to the study of accountability in public administration: provide a basic theory modelling of individual responses to accountability and offer a sophisticated set of tools for establishing causality.

i. A basic theoretical proposition modelling individual behavior in response to accountability: the intuitive politician metaphor. The social contingency model of judgement and choice departs from the assumption that the behavior of individuals is guided by the pursuit of approval and respect from key constituencies (Tetlock, 2002). This search for approval is galvanized by three distinct, yet highly related, motivations, namely the motivation to protect and enhance one's social image in the eyes of relevant others, the motivation to protect and enhance one's self-image, and the motivation to acquire power and resources (Tetlock, 1992 p.338). Thus, following these assumptions, when faced with accountability pressures, individuals will act as *intuitive politicians*, strategically adapting their behavior in order to gain the approval of audiences they consider as important.

Public sector actors as intuitive politicians would thus try to “do well” in their work, since they care about how salient accountability forums will evaluate and perceive them. They want to protect their own personal and professional integrity, and they recognize that being held in high regard will allow them to avoid punishments and gain powers. Their behavior is thus guided by both normative and strategic considerations. What “doing well” would mean in terms of a behavioral response, would be shaped to a large extent by relevant contextual factors, and the expectations set by the accountability forums such as the clients they serve, the professional community they belong to, or their hierarchical principals.

The multifaceted basis of the motivational drive, as well as the ability to account for contextual factors in anticipating the behavioral responses to accountability, are what makes the intuitive politician metaphor a strong and useful theoretical tool. It allows for a more complex and nuanced picture of individual behavior than some of the theoretical tools used in public accountability research, such as agency theory and stewardship theory, which champion single motivational propositions and dominant behaviors. Thus, instead of depicting individuals as exclusively self-interested, like in the case of agency theory, or intrinsically motivated, like in the case of stewardship theory (Dicke, 2002; Schillemans & Bjurström, 2019), the intuitive politician allows for both types of motivations to be accounted for.

Many of the propositions of the intuitive politician metaphor are aligned with the recently developed reputational perspective on accountability (Busuioc & Lodge, 2016, 2017), which in some way testifies to its attractiveness. However, while the reputational perspective on accountability mostly focuses on explaining organizational behavior (Busuioc & Lodge, 2016, p.249, 2017, p.3), the intuitive politician metaphor is centered on the individual. In addition, its firm grounding in behavioral, and more specifically socio-psychological research, makes the

intuitive politician metaphor a particularly suitable tool for investigating individual behavior under accountability pressure.

ii. Sophisticated tool-box for establishing causal effects. Experiments have found their place in the methodological tool-box in many areas of inquiry in public administration in recent years (Bouwman & Grimmelikhuijsen, 2016; James et al., 2017). They are seen as a particularly useful tool for overcoming endogeneity issues often present in social science empirical research, and for establishing clear chains of cause and effect (Blom-Hansen et al., 2015; Bouwman & Grimmelikhuijsen, 2016). Despite their popularity in other domains of public administration, they still appear very rarely in the study of public accountability (for exceptions Bouwman et al., 2018; Ossege, 2012; Schillemans & Giesen, 2020). In fact, the empirical study of public accountability is still to a large extent qualitative and descriptive (Brandsma, 2014; Schillemans & Busuioc, 2014; Yang, 2012).

In contrast to the research on accountability in public administration, the study of accountability in the behavioral sciences has been predominantly carried out with the help of experimental designs (Lerner & Tetlock, 1999). These designs present indispensable guides in the development of a behavioral and causal research agenda for the study of accountability in public administration. They offer practical instructions on how to design experimental tools to investigate the behavioral effects of accountability in terms of manipulations and scenarios, but also whole experimental designs which can be directly replicated with participants from the public sector. In addition, these experiments display how the different elements of accountability mechanisms, namely accountable actors, forums, standards and motivations for account-giving, can be usefully disentangled and isolated and subjected to systematic counterfactual analysis. Thus, the empirical work conducted in the behavioral sciences eases the adoption of experimental methods in the study

of public accountability, and thereby clears the way for systematic causal research of the effects of accountability mechanisms. At the same time, we should be careful not to mindlessly translate experimental designs from the behavioral sciences to public administration, as research findings do not travel unchallenged from the lab to the field, as discussed earlier. This will be explored in more detail in the next section.

Where we need public administration: Professional repertoires and experimental contextualism

The results of the experimental studies presented in this dissertation did not always support the hypotheses outlined by the models derived from the behavioral sciences. In addition, the experimental tools from the behavioral sciences did not always fit the public sector context. This suggests that the discipline of public administration should offer its unique contribution towards the understanding of the behavioral effects of accountability in the public sector, both in theoretical and methodological terms.

i. Accountability management of professionals. The contexts in which accountability mechanisms are enacted in the public sector differ in some important ways from the contexts in which accountability is commonly explored in the behavioral sciences. These discrepancies in the investigated context could account for important differences in observed decision-making behavior between civil servants on the one hand, and the subjects in the experimental behavioral research in accountability on the other hand. Two contextual factors appear as most important here, namely, the exposure to accountability pressures, and the decision stakes.

Public sector professionals are continuously faced with accountability in their professional capacity. At minimum, they are aware that they might be called to give account for their work at

some point, while often, the account-giving presents a routinized element of their work (Hwang & Han, 2017; Schillemans, 2015). Public sector professionals are thus likely to be more accustomed to dealing with accountability pressures than the participants commonly taking part in the accountability studies in the behavioral sciences, who as described in Chapter 2, are commonly undergraduate students. For example, the majority of regulators surveyed in Chapter 5, as well as the civil servants surveyed in Chapter 4, indicated that they regularly face situations of multiple accountability pressures. Dealing with multiple accountability demands is, therefore, a constituting element of their work. This stands in sharp contrast to the “one-off” accountability situations modeled in the experimental behavioral research on accountability, where the accountability pressure is often unanticipated, and therefore likely experienced as much more disruptive. This could potentially be the key reason why we observe differences in decision-making effort as a result of variation in accountability pressure among participants in behavioral studies on accountability, but not among the public sector professionals surveyed in this dissertation. It might take a bit more than routine account-giving to faze public sector professionals.

The professional context in which public sector professionals operate likely matters also in a different way: the decisions these professionals routinely take have arguably higher stakes than the ones modeled in generic behavioral experiments. Many decisions public sector professionals take have far reaching consequences: inspectors can decide the fate of a business; social caseworkers can decide on granting life-changing financial benefits; and middle managers can decide whether to communicate information of crucial importance to the success of a project to higher-ups, or not. The decisions of public sector professionals are not only significant for their clients, hierarchical principals, or the broader public, but also for themselves, as they are professionals, who likely value being seen as competent, are intrinsically motivated to perform their tasks, and might see

their work as means to achieving some societal, personal and strategic interests (Crewson, 1997; Busuioc & Lodge, 2016; Schillemans & Bjurstrøm, 2019). Thus, there are many reasons to expect that public sector professionals would take their decisions seriously, even when they are not under direct pressure of an accountability forum to do so. This is evidently not the case for participants in generic behavioral studies, as their decisions have rarely any significant stakes beyond the prospect of some social embarrassment, or short-term reputational damage.

The ways in which these two contextual factors could affect the behavioral consequences of accountability we observe among public sector professionals is discussed further in the final section of this chapter.

ii. Experimenting with public accountability. Replicating existing experimental designs from the behavioral sciences with public sector actors as participants may appear to be simple enough, with the only hurdle to overcome being the recruitment of these participants. However, when we try to take a step towards greater experimental realism—which is perhaps more highly valued in the field of public administration than in other fields—a whole new set of considerations and challenges emerge. Two such issues are central: the administering of the accountability manipulation, and the consideration of the professional rules and practices in which public sector actors operate.

In the behavioral sciences, the accountability manipulation is commonly administered as the introduction of a requirement to provide a justification for one's decisions, actions or beliefs (Lerner & Tetlock, 1999, p. 255), usually directly to the significant other, i.e. the accountability forum. The accountability forum is thus often either physically present during the account-giving process, or the participants attach information regarding their identity to their responses, so that their work could be linked to them personally and can be evaluated by a (known) forum at a later point.

None of these two modes of introducing the accountability forum is easily applicable to public sector experiments on accountability with higher realism. Assuring the presence of the relevant accountability forums in an experimental setting is not only challenging from a recruitment point of view, it is also difficult in terms of defining adequate representatives of the various forums public sector actors face in reality, such as the professional community, the clients, or the public. While recording the identity of the respondents so that their responses can be submitted to relevant accountability forums for evaluation and linked back to them personally, appears to be a more practically feasible solution, it is against today's regulations and practices in experimental research to do so. Thus, public administration researchers seeking to experimentally examine public accountability are left with having to accept weaker modes of direct administration of the accountability manipulation, where they act themselves as the accountability forums, as in the experimental study presented in Chapter 3. Alternatively, they can seek other, indirect, ways of administering the accountability manipulation, where the accountability forums are merely invoked, as opposed to directly involved, like in the experimental studies presented in Chapters 4 and 5 of this dissertation.

The second challenge comes from the need to understand the context in which public sector actors operate. This pertains to the key rules and practices in their work, as well as the accountability forums they face and the nature of the relationship between the two. In order to construct an experimental task or scenario which meaningfully captures the accountability processes public sector actors are a part of, the researcher has to have a good understanding of the real dilemmas public sector actors face, and the range of actions they can take. This is also key so that participants perceive the experiment as meaningful, and as intended. The large body of qualitative descriptive research on public accountability could be useful in this regard, however, in order to develop

experimental tools which would be tailor-made to the circumstances of an organization, or domain of the public sector, a close collaboration with practitioners in development of the experiment might be necessary. Such a collaboration with practitioners was adopted in the development of the experiments presented in Chapters 4 and 5 of this dissertation.

6.5. Towards a behavioral theory of public accountability?

A complete understanding of the behavioral effects of accountability in the public sector is far from achieved. There are still numerous questions which remain to be examined in more detail. But before scholars continue probing into the workings of accountability mechanisms in various decision-making areas and domains in the public sector, it might be useful to capitalize on the existing knowledge on the operation of accountability mechanisms from both the behavioral sciences and public administration, and develop a preliminary behavioral theory of public sector accountability which could guide future research. The preceding discussions highlighted that the social contingency model of judgement and choice would need to be usefully refined to account for the specific circumstances in which accountability mechanisms in the public sector are enacted. Specifically, two contextual factors appear to be crucial to account for in developing a public sector-specific behavioral model of accountability. These pertain to the application of the accountability pressure, and the stakes in which the decisions in the public sector are made. The ways in which these factors could give rise to unique decision-making behaviors among public sector professionals is discussed in more detail below.

Continuous accountability pressure. Very few civil servants in modern contemporary bureaucracies would be completely blindsided by an expectation to give account for their work.

While some might fall through the cracks and find themselves in the comfort of the ambivalence of their accountability forums (Benjamin & Posner, 2018; Schillemans & Busuioc, 2014), most would have at least an implicit expectation that someone, sometimes, would hold them to account, be it their hierarchical principal, the professional community, or the citizens they serve.

For many civil servants, the management of accountability expectations appears to be a routine aspect of their work. Using tested strategies, coping mechanisms, and prioritization tactics, these civil servants are accustomed to addressing the stream of accountability demands and dilemmas regularly coming their way (Hwang & Han, 2017; Schillemans, 2015). This is not to say that accountability mechanisms do not introduce complexities in their work, because they certainly do, but those complexities rarely come as a surprise to the experienced civil servant.

When account-giving is largely a routinized affair, responding to the demands of an accountability forum is not likely to elicit a great emotional, cognitive, or even behavioral response from the civil servant. The demand would be addressed by accessing the repertoire of tried-and-tested actions, without greatly disturbing the normal workflow. In other words, the public sector actors will use an “acceptability heuristic” (Tetlock, 1992) to respond to a familiar accountability demand, which they anticipate will be satisfactory to the accountability forum.

There are certain situations, however, when public sector professionals are unable to respond to accountability pressures in a routinized fashion. These can appear in new, rapidly changing or evolving organizations, where the operation of accountability mechanisms did not have a chance to settle in just yet (Olsen, 2013). They can also be encountered after changes in the institutional, legal, social and political landscape, when for example a reform broadens the mandate of a public sector organization, a new government announces a radical shift in priorities, or the routine practices of account giving and holding fall apart as they uncover malpractices and problems.

When there is no available acceptability heuristic to employ, the account-giving will present itself as a greater hurdle to the public sector actor, and thus bring about greater behavioral adaptation. Providing an answer to the accountability demand would require greater preparations, anticipations of the possible objections of the accountability forum, and thus more careful consideration. Or in other words, public sector professionals will engage in “preemptive self-criticism” (Tetlock, 1992) and thus weigh the pros and cons of multiple options before they settle on a decision or an action. In these situations, resorting to decision- and blame-avoidance strategies would also become more likely (Hood, 2010). As the decision-making context is characterized with greater uncertainty, public sector actors would be less willing to stick their necks out.

High(er) stakes. The types of decisions public sector professionals regularly face are conceivably more substantive than the types of decisions we encounter in experimental behavioral research on accountability. Public sector professionals make decisions in which they themselves, as well as their accountability forums, have vested interests. They also have their professional reputation to guard and their strategic interests to pursue. They are thus likely to invest more time and effort in the crafting of their responses than an individual who is asked, for example, to express a political opinion to an unfamiliar audience, a decision-making context often encountered in the behavioral experimental research on accountability (like in Green et al., 2000; Tetlock, 1983). This could have several implications on how accountability pressures might affect their behavioral responses.

First, if we consider that the decisions of public sector professionals can bear significant consequences, both for relevant external stakeholders and for the professionals making them, it is likely that they will receive some degree of consideration, even without being prompted by the threat of accountability. In other words, the baseline effort that public sector professionals invest in decisions belonging to their professional domain is likely higher than the baseline effort

decision-makers in generic behavioral studies would invest. As a result, the gains from exerting accountability pressures in terms of decision-making effort among public sector professionals would be smaller than among generic decision-makers.

Second, as public sector professionals often have their own preferences and strategic considerations when it comes to the decisions they make, they would be less inclined to simply follow the explicitly communicated demands or expectations of accountability forums in an effort to gain their approval, as the social-psychological research finds (e.g. Tetlock et al., 1989). Instead they employ more sophisticated strategies to both satisfy their forum, and incorporate their own interests.

To conclude, the preceding discussion offers several propositions and hypotheses which could help explain the unique dynamics of the behavioral responses to accountability pressures in the public sector, and the way they differ from other contexts. Their validity however, should be put to the empirical test. As accountability mechanisms are ascribed both outcomes of improved and worsened performance, it is imperative to understand the causal chains leading to each of those. This dissertation has taken some important steps in that direction, however, much more remains to be done until full understanding of the effects and consequences of accountability in public administration is achieved.

References

- ABDTOPConsult (2019). *Regels en Ruimte Verkenning Maatwerk in dienstverlening en discretionaire ruimte*.
<https://www.rijksoverheid.nl/documenten/rapporten/2020/01/16/bijlage-rapport-abdtopconsult-maatwerk-dienstverlening>
- Aberbach, J. D., & Rockman, B. A. (1997). Bureaucracy: Control, responsiveness, performance. In A. I. Baaklini & H. Desfosses (Eds.), *Designs for democratic stability: Studies in viable constitutionalism* (pp. 73-94). Routledge.
- Acar, M., Guo, C., & Yang, K. (2008). Accountability when hierarchical authority is absent: views from public-private partnership practitioners. *The American Review of Public Administration*, 38(1), 3-23.
- 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.
- Anderson, C. J. (2003). The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological bulletin*, 129(1), 139-167.
- Andeweg, R. B., & Irwin, G. A. (2014). *Governance and Politics of the Netherlands* (4th ed.). Macmillan International Higher Education.
- Andrews, R., Boyne, G. A., Law, J., & Walker, R. M. (2008). Organizational strategy, external regulation and public service performance. *Public administration*, 86(1), 185-203.
- Arkes, H. R., González-Vallejo, C., Bonham, A. J., Kung, Y. H., & Bailey, N. (2010). Assessing the merits and faults of holistic and disaggregated judgments. *Journal of Behavioral Decision Making*, 23(3), 250-270.
- Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*, 52(2), 292-306.
- Ashworth, R., Boyne, G. A., & Walker, R. M. (2002). Regulatory problems in the public sector: theories and cases. *Policy & Politics*, 30(2), 195-211.
- Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, 10(2), 150-161.
- Bach, T., van Thiel, S., Hammerschmid, G., & Steiner, R. (2017). Administrative tradition and management reforms: a comparison of agency chief executive accountability in four Continental Rechtsstaat countries. *Public Management Review*, 19(6), 765-784.
- Bagley, P. L. (2010). Negative affect: A consequence of multiple accountabilities in auditing. *Auditing: A Journal of Practice & Theory*, 29(2), 141-157.

- Barreto, M., & Ellemers, N. (2000). You can't always do what you want: Social identity and self-presentational determinants of the choice to work for a low-status group. *Personality and Social Psychology Bulletin*, 26(8), 891-906.
- Behn, R. D. (2001). *Rethinking democratic accountability*. Brookings Institution Press.
- Benjamin, L. M., & Posner, P. L. (2018). Tax expenditures and accountability: The case of the ambivalent principals. *Journal of Public Administration Research and Theory*, 28(4), 569-582.
- Bevan, G., & Hood, C. (2006). What's measured is what matters: targets and gaming in the English public health care system. *Public Administration*, 84(3), 517-538.
- Biela, J., & Papadopoulos, Y. (2014). The empirical assessment of agency accountability: a regime approach and an application to the German Bundesnetzagentur. *International Review of Administrative Sciences*, 80(2), 362-381.
- Blom-Hansen, J., Morton, R., & Serritzlew, S. (2015). Experiments in public management research. *International Public Management Journal*, 18(2), 151-170.
- Bodenhausen, G. V., Kramer, G. P., & Süsler, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of personality and social psychology*, 66(4), 621-632.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*, 1(2), 97-111.
- Boulianne, S. & Basson, D. (2008). Topic saliency. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 892-892). Sage.
- Bouwman, R., & Grimmelikhuisen, S. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management*, 29(2), 110-131.
- Bouwman, R., Van Thiel, S., van Deemen, A., & Rouwette, E. (2018). Accountability and coalitions: Evidence from a negotiation experiment. *Public Administration Review*, 78(1), 37-47.
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European law journal*, 13(4), 447-468.
- Bovens, M. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *West European Politics*, 33(5), 946-967.

- Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.). (2014). *The Oxford handbook public accountability*. Oxford University Press.
- Bovens, M., Schillemans, T., & Hart, P. T. (2008). Does public accountability work? An assessment tool. *Public administration*, 86(1), 225-242.
- Bovens, M., Schillemans, T., & Goodin, R. E. (2014). Public accountability. *The Oxford handbook of public accountability*, 1(1), 1-22.
- Boyne, G., Day, P., & Walker, R. (2002). The evaluation of public service inspection: A theoretical framework. *Urban Studies*, 39(7), 1197-1212.
- Bozeman, B., & Scott, P. (1992). Laboratory experiments in public policy and management. *Journal of Public Administration Research and Theory*, 2(3), 293-313.
- Braithwaite, J. (1997). On speaking softly and carrying big sticks: Neglected dimensions of a republication separation of powers. *The University of Toronto Law Journal*, 47(3), 305-361.
- Brandsma, G. J. (2014). Quantitative analysis. *The oxford handbook of public accountability*. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 143-158). Oxford University Press.
- Braun, C. (2012). The captive or the broker? Explaining public agency–interest group interactions. *Governance*, 25(2), 291-314.
- Brennan, G. (1999). Institutionalising accountability: A commentary. *Australian Journal of Public Administration*, 58(1), 94-97.
- Bringselius, L. (2017). *Tillitsbaserad styrning och ledning: Ett ramverk*. (Samtal om tillit i styrning). Tillitsdelegationen. https://portal.research.lu.se/portal/files/35941881/Rapport_Tillitsdelegationen_2017_2_u_pplaga_2.pdf
- Brodkin, E. Z. (1997). Inside the welfare contract: Discretion and accountability in state welfare administration. *Social Service Review*, 71(1), 1-33.
- Brodkin, E. Z. (2008). Accountability in street-level organizations. *International Journal of Public Administration*, 31(3), 317-336.
- Brummel, L. (2021). Social Accountability Between Consensus and Confrontation: Developing a Theoretical Framework for Societal Accountability Relationships of Public Sector Organizations. *Administration & Society*.
- Busuioc, M. (2009). Accountability, control and independence: the case of European agencies. *European Law Journal*, 15(5), 599-615.

- Busuioc, E. M., & Lodge, M. (2016). The reputational basis of public accountability. *Governance*, 29(2), 247-263.
- Busuioc, M., & Lodge, M. (2017). Reputation and accountability relationships: Managing accountability expectations through reputation. *Public Administration Review*, 77(1), 91-100.
- Butterfield, R., Edwards, C., & Woodall, J. (2005). The new public management and managerial roles: The case of the police sergeant. *British Journal of Management*, 16(4), 329-341.
- Card, N. A. (2015). *Applied meta-analysis for social science research*. Guilford Publications.
- Carpenter, D. (2010). *Reputation and power: Organizational image and pharmaceutical regulation at the FDA*. Princeton University Press.
- Carpenter, D. P., & Krause, G. A. (2012). Reputation and public administration. *Public Administration Review*, 72(1), 26-32.
- Caseley, J. (2006). Multiple accountability relationships and improved service delivery performance in Hyderabad City, Southern India. *International Review of Administrative Sciences*, 72(4), 531-546.
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, M. H. (2020). *pwr*. (R package version 1.3.0).
- 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.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.
- Cohen, J. R., & Trompeter, G. M. (1998). An examination of factors affecting audit practice development. *Contemporary Accounting Research*, 15(4), 481-504.
- Conway, L. G., Conway, K. R., Gornick, L. J., & Houck, S. C. (2014). Automated integrative complexity. *Political Psychology*, 35(5), 603-624.
- Conway, L. G., Conway, K. R., & Houck, S. C. (2020). Validating Automated Integrative Complexity: Natural Language Processing and the Donald Trump Test. *Journal of Social and Political Psychology*, 8(2), 504-524.
- Coletti, A. L., Sedatole, K. L., & Towry, K. L. (2005). The effect of control systems on trust and cooperation in collaborative environments. *The Accounting Review*, 80(2), 477-500.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15-18.

- Crewson, P. E. (1997). Public-service motivation: Building empirical evidence of incidence and effect. *Journal of public administration research and theory*, 7(4), 499-518.
- Dahlström, C., Teorell, J., Dahlberg, S., Hartmann, F., & Lindberg, A. (2015). *The QoG Expert Survey II Report* (Working paper series 2015:9).
- Daly, A. J. (2009). Rigid response in an age of accountability: The potential of leadership and trust. *Educational Administration Quarterly*, 45(2), 168-216.
- Davis, W. D., Mero, N., & Goodman, J. M. (2007). The interactive effects of goal orientation and accountability on task performance. *Human Performance*, 20(1), 1-21.
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). Toward a stewardship theory of management. *Academy of Management review*, 22(1), 20-47.
- De Dreu, C. K., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of personality and social psychology*, 90(6), 927-943.
- De Dreu, C. K., & van Knippenberg, D. (2005). The possessive self as a barrier to conflict resolution: effects of mere ownership, process accountability, and self-concept clarity on competitive cognitions and behavior. *Journal of personality and social psychology*, 89(3), 345-357.
- De Dreu, C. K., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: a motivated information-processing approach to integrative negotiation. *Journal of personality and social psychology*, 79(6), 975-987.
- Demir, T., Reddick, C. G., Ponomariov, B., & Flink, C. M. (2019). Does Power Corrupt? An Empirical Study of Power, Accountability, and Performance Triangle in Public Administration. *International Journal of Public Administration*, 42(9), 723-740.
- DeZoort, T., Harrison, P., & Taylor, M. (2006). Accountability and auditors' materiality judgments: The effects of differential pressure strength on conservatism, variability, and effort. *Accounting, Organizations and Society*, 31(4-5), 373-390.
- Dicke, L. A. (2002). Ensuring accountability in human services contracting: Can stewardship theory fill the bill?. *The American Review of Public Administration*, 32(4), 455-470.
- Donaldson, L. (2001). *The contingency theory of organizations*. Sage.
- Dubnick, M. J. (2002, August/September). Seeking salvation for accountability [Paper presentation]. *Annual Meeting of the American Political Science Association 2002*, Boston, United States.
- Dubnick, M. J. (2003). Accountability and ethics: Reconsidering the relationships. *International Journal of Organization Theory and Behavior*, 6(3), 405-441.

- Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, 28(3), 376-417.
- Dubnick, M. J., & Frederickson, H. G. (Eds.). (2011). *Accountable governance: Problems and promises*. Routledge.
- Dudley, S. E., & Xie, Z. (2019). Nudging the nudger: Toward a choice architecture for regulators. *Regulation & Governance*.
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463.
- Ewert, B. (2020). Focusing on quality care rather than ‘checking boxes’: How to exit the labyrinth of multiple accountabilities in hybrid healthcare arrangements. *Public Administration*, 98(2), 308-324.
- Feitsma, J. N. P. (2018). The behavioural state: critical observations on technocracy and psychocracy. *Policy sciences*, 51(3), 387-410.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychological methods*, 6(2), 161-180.
- Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, 63(3), 665-694.
- Flinders, M. (2011). Daring to be a Daniel: The pathology of politicized accountability in a monitory democracy. *Administration & Society*, 43(5), 595-619.
- Flinders, M. (2014). The future and relevance of accountability studies. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 661-672). Oxford University Press.
- Flinders, M., & Tonkiss, K. (2016). From ‘poor parenting’ to micro-management: coalition governance and the sponsorship of arm’s-length bodies in the United Kingdom, 2010–13. *International Review of Administrative Sciences*, 82(3), 490-515.
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly*, 59(3), 237-254.
- Frink, D. D., & Ferris, G. R. (1999). The moderating effects of accountability on the conscientiousness-performance relationship. *Journal of Business and Psychology*, 13(4), 515-524.
- Gailmard, S. (2014). Accountability and Principal-Agent Models. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability* (pp. 90-106). Oxford University Press.

- Gilad, S., & Yogev, T. (2012). How reputation regulates regulators: Illustrations from the regulation of retail finance. In T.G. Pollock & M.L. Barnett (Eds.), *The Oxford handbook of corporate reputation* (pp. 321-340). Oxford University Press.
- Green, M. C., Visser, P. S., & Tetlock, P. E. (2000). Coping with accountability cross-pressures: Low-effort evasive tactics and high-effort quests for complex compromises. *Personality and Social Psychology Bulletin*, 26(11), 1380-1391.
- Green, P. E., & Rao, V. R. (1971). Conjoint measurement-for quantifying judgmental data. *Journal of Marketing research*, 8(3), 355-363.
- Gregory, R. (1998). Political responsibility for bureaucratic incompetence: tragedy at Cave Creek. *Public Administration*, 76(3), 519-538.
- Greiling, D., & Spraul, K. (2010). Accountability and the Challenges of Information Disclosure. *Public Administration Quarterly*, 34(3), 338-377.
- 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.
- Grimmelikhuijsen, S. G., & Meijer, A. J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration Research and Theory*, 24(1), 137-157.
- Haccoun, R. R., & Klimoski, R. J. (1975). Negotiator status and accountability source: A study of negotiator behavior. *Organizational Behavior and Human Performance*, 14(3), 342-359.
- Hagafors, R., & Brehmer, B. (1983). Does having to justify one's judgments change the nature of the judgment process?. *Organizational Behavior and Human Performance*, 31(2), 223-232.
- Hainmueller, J., & Hopkins, D. J. (2015). The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science*, 59(3), 529-548.
- Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political analysis*, 22(1), 1-30.
- Halachmi, A. (2002). Performance measurement, accountability, and improved performance. *Public Performance & Management Review*, 25(4), 370-374.
- Halachmi, A. (2014). Accountability overloads. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 560-573). Oxford University Press.

- Hall, A. T., Frink, D. D., & Buckley, M. R. (2017). An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior*, 38(2), 204-224.
- Han, Y., & Perry, J. L. (2020a). Conceptual bases of employee accountability: A psychological approach. *Perspectives on Public Management and Governance*.
- Han, Y., & Perry, J. L. (2020b). Employee accountability: development of a multidimensional scale. *International Public Management Journal*, 23(2), 224-251.
- Harari, M. B., & Rudolph, C. W. (2017). The effect of rater accountability on performance ratings: A meta-analytic review. *Human Resource Management Review*, 27(1), 121-133.
- Harrison, S., & Dowswell, G. (2002). Autonomy and bureaucratic accountability in primary care: what English general practitioners say. *Sociology of Health & Illness*, 24(2), 208-226.
- Hasler, K., Kuebler, D., & Marcinkowski, F. (2016). Over-responsibilised and over-blamed: elected actors in media reporting on network governance. A comparative analysis in eight European metropolitan areas. *Policy & Politics*, 44(1), 135-152.
- Heimann, C. L. (1993). Understanding the Challenger disaster: Organizational structure and the design of reliable systems. *American Political Science Review*, 421-435.
- Hess, T. M., Rosenberg, D. C., & Waters, S. J. (2001). Motivation and representational processes in adulthood: the effects of social accountability and information relevance. *Psychology and aging*, 16(4), 629-642.
- Hill, C. J., & Lynn, L. E. (2004). Is hierarchical governance in decline? Evidence from empirical research. *Journal of public administration research and theory*, 15(2), 173-195.
- Hill, M., & Varone, F. (2014). *The public policy process*. Routledge.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1-28.
- Hoecht, A. (2006). Quality assurance in UK higher education: Issues of trust, control, professional autonomy and accountability. *Higher education*, 51(4), 541-563.
- Hollibaugh Jr, G. E., Miles, M. R., & Newswander, C. B. (2020). Why public employees rebel: guerrilla government in the public sector. *Public Administration Review*, 80(1), 64-74.
- Holt, J. E., Kinchin, G., & Clarke, G. (2012). Effects of peer-assessed feedback, goal setting and a group contingency on performance and learning by 10–12-year-old academy soccer players. *Physical Education & Sport Pedagogy*, 17(3), 231-250.
- Hood, C. (2006). Gaming in targetworld: The targets approach to managing British public services. *Public Administration Review*, 66(4), 515-521.

- Hood, C. (2010). *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton University Press.
- Hood, C., James, O., Scott, C., Jones, G. W., & Travers, T. (1999). *Regulation inside government: Waste watchers, quality police, and sleaze-busters*. Oxford University Press.
- Houck, S. C., Conway III, L. G., & Gornick, L. J. (2014). Automated integrative complexity: Current challenges and future directions. *Political Psychology, 35*(5), 647-659.
- Huber, O., Bär, A. S., & Huber, O. W. (2009). Justification pressure in risky decision making: Search for risk defusing operators. *Acta Psychologica, 130*(1), 17-24.
- Huneke, M. E., Cole, C., & Levin, I. P. (2004). How varying levels of knowledge and motivation affect search and confidence during consideration and choice. *Marketing Letters, 15*(2-3), 67-79.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Sage.
- Hupe, P., & Hill, M. (2007). Street-Level bureaucracy and public accountability. *Public Administration, 85*(2), 279-299.
- Hwang, K. (2016). Accountability practices in public child welfare services. *International Journal of Public Administration, 39*(8), 587-596.
- Hwang, K., & Han, Y. (2017). Public caseworkers' strategies coping with accountability demands. *Journal of Public Affairs, 17*(4), e1669.
- James, O., Jilke, S., & Van Ryzin, G. G. (2017). Causal inference and the design and analysis of experiments. In O. James, S. R. Jilke, & G. G. Van Ryzin (Eds.), *Experiments in public management research: Challenges and contributions* (pp. 59-88). Cambridge University Press.
- James, O., Olsen, A. L., Moynihan, D., & Van Ryzin, G. G. (2020). *Behavioral public performance: How people make sense of government metrics*. Cambridge University Press.
- Jantz, B., Klenk, T., Larsen, F., & Wiggan, J. (2018). Marketization and varieties of accountability relationships in employment services: Comparing Denmark, Germany, and Great Britain. *Administration & Society, 50*(3), 321-345.
- Jarvis, M. D. (2014). Hierarchical Accountability. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 405-420). Oxford University Press.
- Jermias, J. (2006). The influence of accountability on overconfidence and resistance to change: A research framework and experimental evidence. *Management Accounting Research, 17*(4), 370-388.

- 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.
- Jilke, S., & Van Ryzin, G. (2017). Survey Experiments for Public Management Research. In O. James, S. Jilke, & G. G. Van Ryzin (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 117-138). Cambridge University Press.
- Johnson, V. E., & Kaplan, S. E. (1991). Experimental-evidence on the effects of accountability on auditor judgments. *Auditing-A Journal of Practice & Theory*, 10, 96-107.
- Jos, P. H., & Tompkins, M. E. (2004). The accountability paradox in an age of reinvention: The perennial problem of preserving character and judgment. *Administration & Society*, 36(3), 255-281.
- Kader, G. D., & Perry, M. (2007). Variability for categorical variables. *Journal of Statistics Education*, 15(2).
- Karsten, N. (2015). Scrutinize me, please! The drivers, manifestations and implications of accountability-seeking behaviour. *Public Administration*, 93(3), 684-699.
- Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy-confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology*, 61(5), 698-707.
- Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, 31(2), 231-245.
- Kim, S. E. (2005). Balancing competing accountability requirements: Challenges in performance improvement of the nonprofit human services agency. *Public Performance & Management Review*, 29(2), 145-163.
- Kim, Y. (2018). Analyzing accountability relationships in a crisis: Lessons from the Fukushima disaster. *The American Review of Public Administration*, 48(7), 743-760.
- Klijn, E. H., & Koppenjan, J. F. (2014). Accountable networks. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 242-257). Oxford University Press.
- Klimoski, R. J. (1972). The effects of intragroup forces on intergroup conflict resolution. *Organizational Behavior and Human Performance*, 8(3), 363-383.
- Klingner, D. E., Nalbandian, J., & Romzek, B. S. (2002). Politics, administration, and markets: Conflicting expectations and accountability. *The American Review of Public Administration*, 32(2), 117-144.
- Koch, C., & Wüstemann, J. (2014). Experimental analysis. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 127-142). Oxford University Press.

- Koonce, L., Anderson, U., & Marchant, G. (1995). Justification of decisions in auditing. *Journal of Accounting Research*, 33(2), 369-384.
- Koop, C. (2014). Theorizing and explaining voluntary accountability. *Public Administration*, 92(3), 565-581.
- Koppell, J. G. (2005). Pathologies of accountability: ICANN and the challenge of “multiple accountabilities disorder”. *Public administration review*, 65(1), 94-108.
- Kroon, M. B. R., t Hart, P., & van Kreveld, D. (1991). Managing Group Decision Making Processes: Individual Versus Collective Accountability and Groupthink. *International Journal of Conflict Management*, 2(2), 91-115.
- Kuhlmann, S., & Wollmann, H. (2019). *Introduction to comparative public administration: Administrative systems and reforms in Europe*. Edward Elgar Publishing.
- Kwon, I. (2014). Motivation, discretion, and corruption. *Journal of Public Administration Research and Theory*, 24(3), 765-794.
- Lægreid, P. (2014). Accountability and new public management. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 324-338). Oxford University Press.
- Lægreid, P., & Neby, S. (2016). Gaming, accountability and trust: DRGs and activity-based funding in Norway. *Financial Accountability & Management*, 32(1), 57-79.
- 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.
- Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. (1999). Motivated search: Effects of choice accountability, issue involvement, and prior knowledge on information acquisition and use. *Journal of Business Research*, 45(1), 75-88.
- Leeper, T. J. (2018). *cregg: Simple Conjoint Analyses and Visualization*. (R package version 0.2.1).
- Leeper, T. J., Hobolt, S. B., & Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis*, 28(2), 207-221.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, 125(2), 255-275.
- Levi-Faur, D. (2005). The global diffusion of regulatory capitalism. *Annals of the American Academy of Political and Social Science*, 598(1), 12-32.
- Lieberherr, E., & Thomann, E. (2019). Street-level bureaucracy research and accountability beyond hierarchy. In P. Hupe (Ed.), *Research handbook on street-level bureaucracy* (pp. 223-239). Edward Elgar Publishing.

- Lierl, M. (2016). Social sanctions and informal accountability: Evidence from a laboratory experiment. *Journal of Theoretical Politics*, 28(1), 74-104.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.
- Lindberg, S. I. (2013). Mapping accountability: core concept and subtypes. *International review of administrative sciences*, 79(2), 202-226.
- Mala, R., Chand, P., & Patel, C. (2018). Influence of experience and accountability on information use and judgments of accountants. *Journal of International Accounting Research*, 17(3), 135-151.
- Mansbridge, J. (2014). A contingency theory of accountability. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 55-68). Oxford University Press.
- 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.
- Mero, N. P., Guidice, R. M., & Brownlee, A. L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management*, 33(2), 223-252.
- Metzl, M., & Speckbacher, G. (2015). How multiple accountabilities and means-end decoupling influence role ambiguity and job satisfaction. *Academy of Management Proceedings*, 2015(1), 18650.
- Mills, R. W., & Koliba, C. J. (2015). The challenge of accountability in complex regulatory networks: The case of the Deepwater Horizon oil spill. *Regulation & Governance*, 9(1), 77-91.
- Mills, R. W., Koliba, C. J., & Reiss, D. R. (2018). Ensuring compliance from 35,000 feet: Accountability and trade-offs in aviation safety regulatory networks. *Administration & Society*, 50(10), 1478-1507.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2020). *Kabinetsreactie Regels en ruimte – Verkenning Maatwerk in dienstverlening en discretionaire ruimte*. <https://www.rijksoverheid.nl/documenten/kamerstukken/2020/01/16/kabinetsreactie-rapport-ruimte-voor-regels-maatwerk-in-dienstverlening>
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, 20(6), 559-572.
- Molander, A., Grimen, H., & Eriksen, E. O. (2012). Professional discretion and accountability in the welfare state. *Journal of Applied Philosophy*, 29(3), 214-230.

- Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. Cambridge University Press.
- Mulgan, R. (2003a). *Holding Power to Account: Accountability in Modern Democracies*. Basingstoke: Palgrave Macmillan.
- Mulgan, R. (2003b). One cheer for hierarchy-Accountability in disjointed governance. *Political science*, 55(2), 6-18.
- Mulgan, R. (2014). Accountability deficits. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 545-559). Oxford University Press.
- Murphy, M., & Skillen, P. (2015). The politics of time on the frontline: Street level bureaucracy, professional judgment, and public accountability. *International Journal of Public Administration*, 38(9), 632-641.
- O'Donnell, G. A. (1998). Horizontal accountability in new democracies. *Journal of democracy*, 9(3), 112-126.
- O'Loughlin, M. G. (1990). What is bureaucratic accountability and how can we measure it?. *Administration & Society*, 22(3), 275-302.
- Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of political Economy*, 115(2), 200-249.
- Olsen, J. P. (2013). The institutional basis of democratic accountability. *West European Politics*, 36(3), 447-473.
- Ordóñez, L. D., Benson III, L., & Beach, L. R. (1999). Testing the compatibility test: How instructions, accountability, and anticipated regret affect prechoice screening of options. *Organizational Behavior and Human Decision Processes*, 78(1), 63-80.
- Ossege, C. (2012). Accountability—are We Better off Without It? An empirical study on the effects of accountability on public managers' work behaviour. *Public Management Review*, 14(5), 585-607.
- Overman, S. (2020). Aligning accountability arrangements for ambiguous goals: the case of museums. *Public Management Review*, 1-21.
- Overman, S., Schillemans, T., & Grimmelikhuijsen, S. (2020). A validated measurement for felt relational accountability in the public sector: gauging the account holder's legitimacy and expertise. *Public Management Review*, 1-20.
- Overman, S., Van Genugten, M., & Van Thiel, S. (2015). Accountability after structural disaggregation: Comparing agency accountability arrangements. *Public Administration*, 93(4), 1102-1120.

- Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22-27.
- Palumbo, A. (2017). *Political accountability*. Routledge.
- Pennington, J., & Schlenker, B. R. (1999). Accountability for consequential decisions: Justifying ethical judgments to audiences. *Personality and Social Psychology Bulletin*, 25(9), 1067-1081.
- Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of public administration research and theory*, 6(1), 5-22.
- Perry, M., & Kader, G. (2005). Variation as unalikeability. *Teaching Statistics*, 27(2), 58-60.
- Peters, B. G. (2014). Accountability in public administration. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 211-225). Oxford University Press.
- Pires, R. R. (2011). Beyond the fear of discretion: Flexibility, performance, and accountability in the management of regulatory bureaucracies. *Regulation & Governance*, 5(1), 43-69.
- Pollitt, C. (2003). *The essential public manager*. Open University Press.
- Pollitt, C., & Hupe, P. (2011). Talking about government: The role of magic concepts. *Public Management Review*, 13(5), 641-658.
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.
- Power, M. (2003). Evaluating the audit explosion. *Law & policy*, 25(3), 185-202.
- Rausch, A., & Brauneis, A. (2015). It's about how the task is set: the inclusion–exclusion effect and accountability in preprocessing management information. *Central European Journal of Operations Research*, 23(2), 313-344.
- Ready, J. T., & Young, J. T. (2015). The impact of on-officer video cameras on police–citizen contacts: Findings from a controlled experiment in Mesa, AZ. *Journal of experimental criminology*, 11(3), 445-458.
- Reynaers, A. M., & Grimmelhuijsen, S. (2015). Transparency in public–private partnerships: Not so bad after all?. *Public Administration*, 93(3), 609-626.
- Roch, S. G. (2006). Discussion and consensus in rater groups: Implications for behavioral and rating accuracy. *Human Performance*, 19(2), 91-115.
- Rock, E. (2020). *Measuring Accountability in Public Governance Regimes*. Cambridge University Press.
- Romzek, B. S. (2000). Dynamics of public sector accountability in an era of reform. *International review of administrative sciences*, 66(1), 21-44.

- Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the public sector: Lessons from the Challenger tragedy. *Public Administration Review*, 47(3), 227-238.
- Romzek, B. S., & Ingraham, P. W. (2000). Cross pressures of accountability: Initiative, command, and failure in the Ron Brown plane crash. *Public Administration Review*, 60(3), 240-253.
- Romzek, B. S., & Johnston, J. M. (2005). State social services contracting: Exploring the determinants of effective contract accountability. *Public Administration Review*, 65(4), 436-449.
- Romzek, B. S., LeRoux, K., & Blackmar, J. M. (2012). A preliminary theory of informal accountability among network organizational actors. *Public Administration Review*, 72(3), 442-453.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3), 638-641.
- Schillemans, T. (2008). Accountability in the shadow of hierarchy: The horizontal accountability of agencies. *Public Organization Review*, 8(2), 175.
- Schillemans, T. (2010). Redundant accountability: The joint impact of horizontal and vertical accountability on autonomous agencies. *Public Administration Quarterly*, 300-337.
- Schillemans, T. (2011). Does horizontal accountability work? Evaluating potential remedies for the accountability deficit of agencies. *Administration & Society*, 43(4), 387-416.
- Schillemans, T. (2012). Double-edged swords: Expert-stakeholders as (slightly) unreliable instruments for control and autonomy of executive agencies. *International Journal of Public Administration*, 35(6), 421-433.
- Schillemans, T. (2013). *The public accountability review. A meta-analysis of public accountability research in six academic disciplines* (Working paper). <https://dspace.library.uu.nl/handle/1874/275784>
- Schillemans, T. (2015). Managing public accountability: How public managers manage public accountability. *International journal of public administration*, 38(6), 433-441.
- Schillemans, T. (2016). Calibrating Public Sector Accountability: Translating experimental findings to public sector accountability. *Public Management Review*, 18(9), 1400-1420.
- Schillemans, T., & Bjurström, K. H. (2019). Trust and verification: balancing agency and stewardship theory in the governance of agencies. *International Public Management Journal*, 1-35.

- Schillemans, T., & Busuioc, M. (2015). Predicting public sector accountability: From agency drift to forum drift. *Journal of Public Administration Research and Theory*, 25(1), 191-215.
- Schillemans, T., & Bovens, M. (2011). The Challenge of Multiple Accountability: Does Redundancy Lead to Overload. In M. J. Dubnick & H. G. Frederickson (Eds.), *Accountable governance: Problems and promises*, (pp.3-21). Routledge.
- Schillemans, T., & Giesen, I. (2020). De kwaliteit van (juridische) oordelen. *Nederlands juristenblad*, 95(3), 184-193.
- Schillemans, T., Overman, S., Fawcett, P., et al. (2020). Understanding Felt Accountability. *Governance*, 1– 24.
- Schoenrade, P. A., Batson, C. D., Brandt, J. R., & Loud, R. E. (1986). Attachment, accountability, and motivation to benefit another not in distress. *Journal of Personality and Social Psychology*, 51(3), 557-563.
- Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, 43(4), 539-552.
- Schul, Y., & Mayo, R. (2003). Searching for certainty in an uncertain world: The difficulty of giving up the experiential for the rational mode of thinking. *Journal of Behavioral Decision Making*, 16(2), 93-106.
- Schwabenland, C., & Hirst, A. (2020). Hybrid accountabilities and managerial agency in the third sector. *Public Administration*, 98(2), 325-339.
- Scott, C. (2000). Accountability in the regulatory state. *Journal of law and society*, 27(1), 38-60.
- Sedikides, C., Herbst, K. C., Hardin, D. P., & Dardis, G. J. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, 83(3), 592-605.
- Shankar, P. G., & Tan, H. T. (2006). Determinants of audit preparers' workpaper justifications. *The Accounting Review*, 81(2), 473-495.
- Shortell, S. M., Waters, T. M., Clarke, K. W., & Budetti, P. P. (1998). Physicians as double agents: maintaining trust in an era of multiple accountabilities. *Jama*, 280(12), 1102-1108.
- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, 65(1), 1-17.
- Simon, H. A. (1947). *Administrative behavior. A study of decision-making processes in administrative organization*. Macmillan.

- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, 51(3), 416-446.
- Simonson, I., & Staw, B. M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of applied psychology*, 77(4), 419-426.
- Sinclair, A. (1995). The chameleon of accountability: Forms and discourses. *Accounting, organizations and Society*, 20(2-3), 219-237.
- Steijn, B., & Noordegraaf, M. (Eds.). (2014). *Professionals under pressure: The reconfiguration of professional work in changing public services*. Amsterdam University Press.
- Strøm, K. (2000). Delegation and accountability in parliamentary democracies. *European journal of political research*, 37(3), 261-290.
- Suedfeld, P. (2010). The cognitive processing of politics and politicians: Archival studies of conceptual and integrative complexity. *Journal of personality*, 78(6), 1669-1702.
- Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: a free and simple tool for meta-analysis. *Research synthesis methods*, 8(4), 537-553.
- Tan, H. T., & Jamal, K. (2006). Managing perceptions of technical competence: How well do auditors know how others view them?. *Contemporary Accounting Research*, 23(3), 761-787.
- Tan, H. T., & Kao, A. (1999). Accountability effects on auditors' performance: The influence of knowledge, problem-solving ability, and task complexity. *Journal of Accounting Research*, 37(1), 209-223.
- Ten Velden, F. S., Beersma, B., & De Dreu, C. K. (2010). It takes one to tango: The effects of dyads' epistemic motivation composition in negotiation. *Personality and Social Psychology Bulletin*, 36(11), 1454-1466.
- Terman, J. N., & Yang, K. (2016). Reconsidering gaming in an accountability relationship: The case of minority purchasing in Florida. *Public Performance & Management Review*, 40(2), 281-309.
- Tetlock, P. E. (1983a). Accountability and complexity of thought. *Journal of personality and social psychology*, 45(1), 74-83.
- Tetlock, P. E. (1983b). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 46(4), 285-292.
- Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, 48(3), 227-236.

- Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in experimental social psychology*, 25, 331-376.
- Tetlock, P. E. (1999). Accountability theory: Mixing properties of human agents with properties of social systems. In J. M. Levine, D. M. Messick, & L. L. Thompson (Eds.), *Shared cognition in organizations: The management of knowledge* (pp.117-137). Psychology Press.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psychological review*, 109(3), 451-471.
- Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of personality and social psychology*, 57(3), 388-398.
- Tetlock, P. E., & Boettger, R. (1994). Accountability amplifies the status quo effect when change creates victims. *Journal of Behavioral Decision Making*, 7(1), 1-23.
- Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of personality and social psychology*, 52(4), 700-709.
- Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of personality and social psychology*, 57(4), 632-640.
- Thomann, E., Hupe, P., & Sager, F. (2018). Serving many masters: Public accountability in private policy implementation. *Governance*, 31(2), 299-319.
- Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology*, 66(3), 474-489.
- Thompson, L. (1995). "They Saw a Negotiation": Partisanship and Involvement. *Journal of Personality and Social Psychology*, 68(5), 839-853.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, 14(3), 207-222.
- Tummers, L. (2020). Behavioral public administration. *Oxford Research Encyclopedia of Politics*.
- Turner, C. W. (2001). Accountability demands and the auditor's evidence search strategy: The influence of reviewer preferences and the nature of the response (belief vs. action). *Journal of Accounting Research*, 39(3), 683-706.
- Van Hiel, A., & Schittekatte, M. (1998). Information exchange in context: Effects of gender composition of group, accountability, and intergroup perception on group decision making. *Journal of Applied Social Psychology*, 28(22), 2049-2067.

- Van Thiel, S., & Yesilkagit, K. (2011). Good neighbours or distant friends? Trust between Dutch ministries and their executive agencies. *Public Management Review*, 13(6), 783-802.
- Verhoest, K., Van Thiel, S., Bouckaert, G., & Lægreid, P. (Eds.). (2012). *Government agencies: Practices and lessons from 30 countries*. Springer.
- Verschuere, B., Verhoest, K., Meyers, F., & Peters, B. G. (2006). Accountability and accountability arrangements in public agencies. In T. Christensen & P. Lægreid (Eds.), *Autonomy and regulation: Coping with agencies in the modern state* (pp. 268-300). Edward Elgar Publishing.
- Von Hippel, P. T. (2015). The heterogeneity statistic I² can be biased in small meta-analyses. *BMC medical research methodology*, 15(1), 1-8.
- Voorn, B., van Genugten, M., & van Thiel, S. (2019). Multiple principals, multiple problems: Implications for effective governance and a research agenda for joint service delivery. *Public Administration*, 97(3), 671-685.
- Walker, R. M., Brewer, G. A., Lee, M. J., Petrovsky, N., & van Witteloostuijn, A. (2018). Best Practice Recommendations for Replicating Experiments in Public Administration. *Journal of Public Administration Research and Theory*.
- Warren, M. E. (2014). Accountability and democracy. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 39-54). Oxford University Press.
- Waterman, R. W., & Meier, K. J. (1998). Principal-agent models: an expansion?. *Journal of public administration research and theory*, 8(2), 173-202.
- Weigold, M. F., & Schlenker, B. R. (1991). Accountability and risk taking. *Personality and Social Psychology Bulletin*, 17(1), 25-29.
- Weldon, E., & Gargano, G. M. (1988). Cognitive loafing: The effects of accountability and shared responsibility on cognitive effort. *Personality and Social Psychology Bulletin*, 14(1), 159-171.
- Willems, T., & Van Dooren, W. (2011). Lost in diffusion? How collaborative arrangements lead to an accountability paradox. *International Review of Administrative Sciences*, 77(3), 505-530.
- Willems, T., & Van Dooren, W. (2012). Coming to terms with accountability: Combining multiple forums and functions. *Public Management Review*, 14(7), 1011-1036.
- Yang, K. (2012). Further understanding accountability in public organizations: Actionable knowledge and the structure–agency duality. *Administration & Society*, 44(3), 255-284.
- Yang, K. (2014). Qualitative analysis. In M. Bovens, R.E. Goodin & T. Schillemans (Eds.), *The Oxford handbook public accountability* (pp. 159-176). Oxford University Press

Appendices

Appendix Chapter 2

2.A. Systematic review search log and PRISMA flow diagram

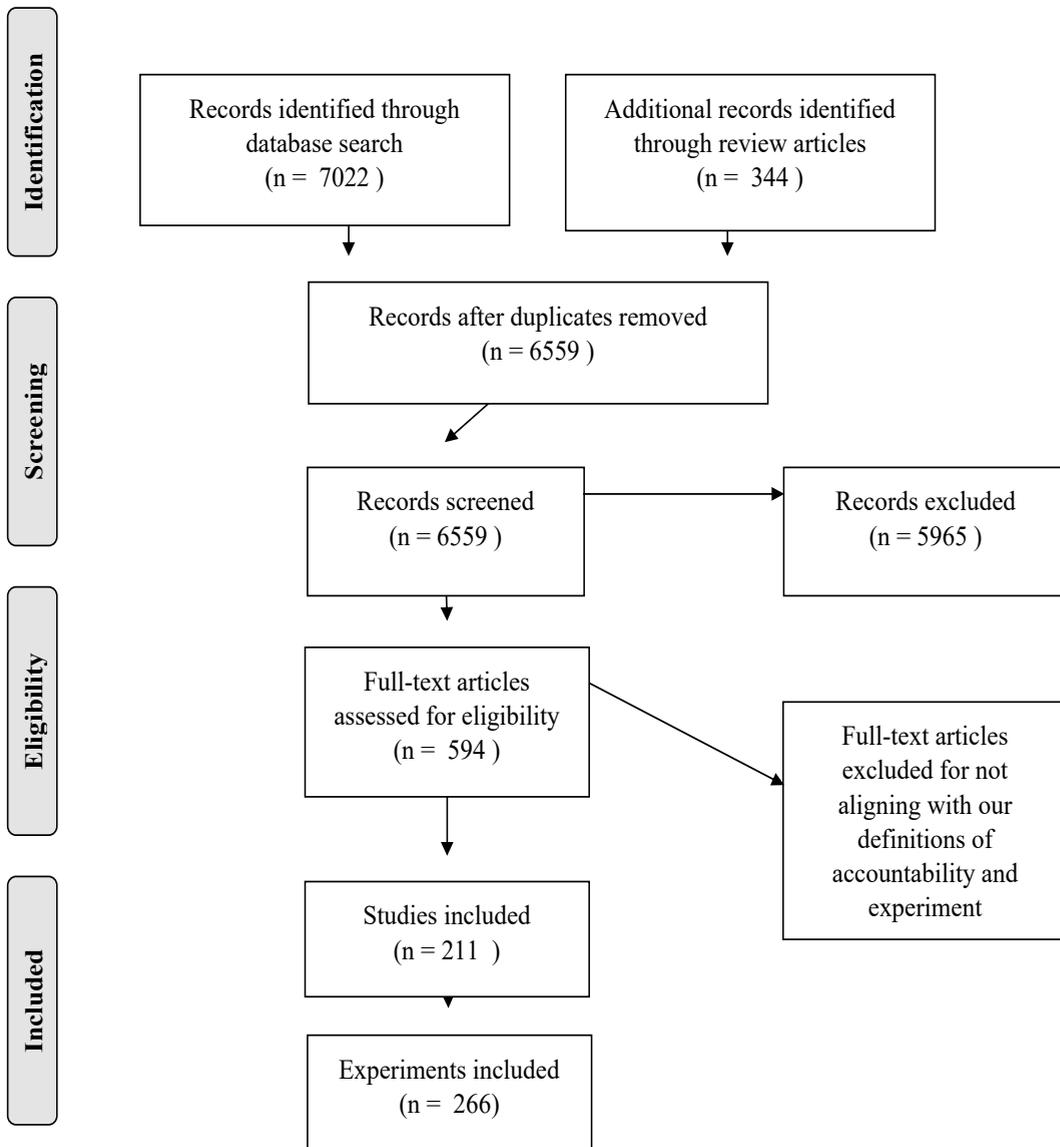
Table 2.A.1 Search log

Database	Date of search	Obtained results	Search log
Science Direct	10.02.2017	101	pub-date > 1822 and pub-date < 2017 and TITLE-ABSTR-KEY(accountability) and TITLE-ABSTR-KEY(experiment)[All Sources(- All Sciences - ,Business, Management and Accounting, Decision Sciences, Economics, Econometrics and Finance, Psychology, Social Sciences)].
Web of Science	13.02.2017	426	(from Web of Science Core Collection) You searched for: TOPIC: (accountability) AND TOPIC: (experiment) Refined by: LANGUAGES: (ENGLISH) AND WEB OF SCIENCE CATEGORIES: (POLITICAL SCIENCE OR PSYCHOLOGY SOCIAL OR ECONOMICS OR BUSINESS FINANCE OR SOCIOLOGY OR MANAGEMENT OR PSYCHOLOGY APPLIED OR PSYCHOLOGY OR BUSINESS OR EDUCATION EDUCATIONAL RESEARCH OR PUBLIC ADMINISTRATION OR LAW OR SOCIAL SCIENCES INTERDISCIPLINARY OR PSYCHOLOGY MULTIDISCIPLINARY OR PSYCHOLOGY EXPERIMENTAL OR EDUCATION SCIENTIFIC DISCIPLINES) AND DOCUMENT TYPES: (ARTICLE OR PROCEEDINGS PAPER OR BOOK CHAPTER) Timespan: 1900-2016. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.
Scopus	13.02.2017	497	TITLE-ABS-KEY ("accountability" AND "experiment") AND PUBYEAR < 2017 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "b")) AND (LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "MULT"))

EBSCO	23.02.2017	350	accountability AND experiment Up to 2016 Language English Default search (not full text)
JSTOR	23.02.2017	5,648	((accountability) AND (experiment)) AND la:(eng OR en) AND disc:(politicalscience-discipline OR health-discipline OR sociology-discipline OR laboremploymentrelations-discipline OR developmentstudies-discipline OR finance-discipline OR education-discipline OR publichealth-discipline OR socialwork-discipline OR peaceconflictstudies-discipline OR law-discipline OR interrela-discipline OR publicpolicy-discipline OR psychology-discipline OR economics-discipline OR manorgbeha-discipline OR markadvert-discipline OR business-discipline OR communicationstudies-discipline)

Note: The searched key words were: “accountability” and “experiment”. The time period was: no start date – 2016. The chosen filed was: behavioral sciences. The selected sources were: published journal articles and books (+ conference papers).

Figure 2.A.1 PRISMA flow diagram



2.B. Complete list of studies included in the review

- Adelberg, S., & Batson, C. D. (1978). Accountability and helping: When needs exceed resources. *Journal of Personality and Social Psychology*, 36(4), 343-350.
- Ahluwalia, R. (2002). How prevalent is the negativity effect in consumer environments?. *Journal of consumer research*, 29(2), 270-279.
- Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., Székely, Á., & Villatoro, D. (2016). Counter-punishment, communication, and cooperation among partners. *Frontiers in behavioral neuroscience*, 10, 53.
- Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology*, 47(2), 349-356.
- Arkes, H. R., González-Vallejo, C., Bonham, A. J., Kung, Y. H., & Bailey, N. (2010). Assessing the merits and faults of holistic and disaggregated judgments. *Journal of Behavioral Decision Making*, 23(3), 250-270.
- Ashton, R. H. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, 28, 148-180.
- Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*, 52(2), 292-306.
- Austin, W., & Walster, E. (1975). Equity with the world: The trans-relational effects of equity and inequity. *Sociometry*, 38(4), 474-496.
- Bardolet, D., Fox, C. R., & Lovallo, D. (2011). Corporate capital allocation: A behavioral perspective. *Strategic Management Journal*, 32(13), 1465-1483.
- Barreto, M., & Ellemers, N. (2000). You can't always do what you want: Social identity and self-presentational determinants of the choice to work for a low-status group. *Personality and Social Psychology Bulletin*, 26(8), 891-906.
- Bartlett, G., Johnson, E., & Reckers, P. (2014). Accountability and role effects in balanced scorecard performance evaluations when strategy timeline is specified. *European Accounting Review*, 23(1), 143-165.
- Bartunek, J. M., Benton, A. A., & Keys, C. B. (1975). Third party intervention and the bargaining behavior of group representatives. *Journal of Conflict Resolution*, 19(3), 532-557.
- Beckner, D., Highhouse, S., & Hazer, J. T. (1998). Effects of upward accountability and rating purpose on peer-rater inflation and delay: a field experiment. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 19(2), 209-214.

- Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of personality and social psychology*, 66(4), 621-632.
- Bolton, L. E. (2003). Stickier priors: The effects of nonanalytic versus analytic thinking in new product forecasting. *Journal of Marketing Research*, 40(1), 65-79.
- Boyle, D. M., DeZoort, F. T., & Hermanson, D. R. (2015). The effects of internal audit report type and reporting relationship on internal auditors' risk judgments. *Accounting Horizons*, 29(3), 695-718.
- Breugh, J. A., & Klimoski, R. J. (1977). The choice of a group spokesman in bargaining: Member or outsider?. *Organizational Behavior and Human Performance*, 19(2), 325-336.
- Brown, C. L. (1999). "Do the Right Thing:" Diverging Effects of Accountability in a Managerial Context. *Marketing Science*, 18(3), 230-246.
- Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology*, 87(1), 185-191.
- Buchman, T. A., Tetlock, P. E., & Reed, R. O. (1996). Accountability and auditors' judgments about contingent events. *Journal of Business Finance & Accounting*, 23(3), 379-398.
- Campbell, B., & Mark, M. M. (2015). How analogue research can advance descriptive evaluation theory: Understanding (and improving) stakeholder dialogue. *American Journal of Evaluation*, 36(2), 204-220.
- Carnevale, P. J., Pruitt, D. G., & Britton, S. D. (1979). Looking tough: The negotiator under constituent surveillance. *Personality and Social Psychology Bulletin*, 5(1), 118-121.
- Chang, C. C., Chuang, S. C., Cheng, Y. H., & Huang, T. Y. (2012). The compromise effect in choosing for others. *Journal of Behavioral Decision Making*, 25(2), 109-122.
- Chang, L. J., Cheng, M. M., & Trotman, K. T. (2013). The effect of outcome and process accountability on customer-supplier negotiations. *Accounting, Organizations and Society*, 38(2), 93-107.
- Chernev, A. (2005). Context effects without a context: attribute balance as a reason for choice. *Journal of Consumer Research*, 32(2), 213-223.
- Chernev, A. (2006). Articulation compatibility in eliciting price bids. *Journal of Consumer Research*, 33(3), 329-341.
- Chernev, A. (2006). Differentiation and parity in assortment pricing. *Journal of Consumer Research*, 33(2), 199-210.
- Clopton, S. W. (1984). Seller and buying firm factors affecting industrial buyers' negotiation behavior and outcomes. *Journal of Marketing Research*, 39-53.

- Cohen, J. R., & Trompeter, G. M. (1998). An examination of factors affecting audit practice development. *Contemporary Accounting Research*, 15(4), 481-504.
- Coletti, A. L., Sedatole, K. L., & Towry, K. L. (2005). The effect of control systems on trust and cooperation in collaborative environments. *The Accounting Review*, 80(2), 477-500.
- Conlon, E. J., & Wolf, G. (1980). The moderating effects of strategy, visibility, and involvement on allocation behavior: An extension of Staw's escalation paradigm. *Organizational Behavior and Human Performance*, 26(2), 172-192.
- Connolly, T., Reb, J., & Kausel, E. E. (2013). Regret salience and accountability in the decoy effect. *Judgment and Decision making*, 8(2), 136-149.
- Corneille, O., Leyens, J. P., Yzerbyt, V. Y., & Walther, E. (1999). Judgeability concerns: the interplay of information, applicability, and accountability in the overattribution bias. *Journal of Personality and Social Psychology*, 76(3), 377-387.
- Cornell, R. M., Eining, M. M., & Hu, P. J. H. (2011). The effects of process accountability on individuals' use of a familiar technology. *Journal of Information Systems*, 25(1), 109-128.
- Cvetkovich, G. (1978). Cognitive accommodation, language, and social responsibility. *Social Psychology*, 41(2), 149-155.
- David, P., Liu, K., & Myser, M. (2004). Methodological artifact or persistent bias? Testing the robustness of the third-person and reverse third-person effects for alcohol messages. *Communication Research*, 31(2), 206-233.
- Davis, W. D., Mero, N., & Goodman, J. M. (2007). The interactive effects of goal orientation and accountability on task performance. *Human Performance*, 20(1), 1-21.
- De Cremer, D., & Van Dijk, E. (2009). Paying for sanctions in social dilemmas: The effects of endowment asymmetry and accountability. *Organizational Behavior and Human Decision Processes*, 109(1), 45-55.
- De Dreu, C. K., & van Knippenberg, D. (2005). The possessive self as a barrier to conflict resolution: Effects of mere ownership, process accountability, and self-concept clarity on competitive cognitions and behavior. *Journal of personality and social psychology*, 89(3), 345-357.
- De Dreu, C. K., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of personality and social psychology*, 90(6), 927-943.
- De Dreu, C. K., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: a motivated information-processing approach to integrative negotiation. *Journal of personality and social psychology*, 79(6), 975-987.

- De Langhe, B., van Osselaer, S. M., & Wierenga, B. (2011). The effects of process and outcome accountability on judgment process and performance. *Organizational Behavior and Human Decision Processes*, *115*(2), 238-252.
- DeAndrea, D. C., Tom Tong, S., Liang, Y. J., Levine, T. R., & Walther, J. B. (2012). When do people misrepresent themselves to others? The effects of social desirability, ground truth, and accountability on deceptive self-presentations. *Journal of Communication*, *62*(3), 400-417.
- DeZoort, F. T., & Harrison, P. D. (2018). Understanding auditors' sense of responsibility for detecting fraud within organizations. *Journal of Business Ethics*, *149*(4), 857-874.
- DeZoort, T., Harrison, P., & Taylor, M. (2006). Accountability and auditors' materiality judgments: The effects of differential pressure strength on conservatism, variability, and effort. *Accounting, Organizations and Society*, *31*(4-5), 373-390.
- Dobbs, M., & Crano, W. D. (2001). Outgroup accountability in the minimal group paradigm: Implications for aversive discrimination and social identity theory. *Personality and Social Psychology Bulletin*, *27*(3), 355-364.
- Druckman, D., Benton, A. A., Ali, F., & Bagur, J. S. (1976). Cultural differences in bargaining behavior: India, Argentina, and the United States. *Journal of Conflict Resolution*, *20*(3), 413-452.
- Fandt, P. M., & Ferris, G. R. (1990). The management of information and impressions: When employees behave opportunistically. *Organizational Behavior and Human Decision Processes*, *45*(1), 140-158.
- Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, *20*(2), 303-315.
- Ford, J. K., & Weldon, E. (1981). Forewarning and accountability: Effects on memory-based interpersonal judgments. *Personality and Social Psychology Bulletin*, *7*(2), 264-268.
- Ford, T. E., Gambino, F., Lee, H., Mayo, E., & Ferguson, M. A. (2004). The role of accountability in suppressing managers' preinterview bias against African-American sales job applicants. *Journal of Personal Selling & Sales Management*, *24*(2), 113-124.
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly*, *59*(3), 237-254.
- Fox, F. V., & Staw, B. M. (1979). The trapped administrator: Effects of job insecurity and policy resistance upon commitment to a course of action. *Administrative Science Quarterly*, *24*(3), 449-471.
- Friedrich, J., Barnes, P., Chapin, K., Dawson, I., Garst, V., & Kerr, D. (1999). Psychophysical numbing: When lives are valued less as the lives at risk increase. *Journal of Consumer Psychology*, *8*(3), 277-299.

- Frink, D. D., & Ferris, G. R. (1998). Accountability, impression management, and goal setting in the performance evaluation process. *Human relations*, 51(10), 1259-1283.
- Frink, D. D., & Ferris, G. R. (1999). The moderating effects of accountability on the conscientiousness-performance relationship. *Journal of Business and Psychology*, 13(4), 515-524.
- Galinsky, A. D., Magee, J. C., Rus, D., Rothman, N. B., & Todd, A. R. (2014). Acceleration with steering: The synergistic benefits of combining power and perspective-taking. *Social Psychological and Personality Science*, 5(6), 627-635.
- Gelfand, M. J., & Realo, A. (1999). Individualism-collectivism and accountability in intergroup negotiations. *Journal of Applied Psychology*, 84(5), 721-736.
- Girvan, E. J., Deason, G., & Borgida, E. (2015). The generalizability of gender bias: Testing the effects of contextual, explicit, and implicit sexism on labor arbitration decisions. *Law and Human Behavior*, 39(5), 525-537.
- Glover, S. M. (1997). The influence of time pressure and accountability on auditors' processing of nondiagnostic information. *Journal of Accounting Research*, 35(2), 213-226.
- Gordon, R. A., & Stuecher, U. (1992). The effect of anonymity and increased accountability on the linguistic complexity of teaching evaluations. *The Journal of psychology*, 126(6), 639-649.
- Gordon, R. A., Rozelle, R. M., & Baxter, J. C. (1988). The effect of applicant age, job level, and accountability on the evaluation of job applicants. *Organizational Behavior and Human Decision Processes*, 41(1), 20-33.
- Gordon, R. A., Rozelle, R. M., & Baxter, J. C. (1989). The effect of applicant age, job level, and accountability on perceptions of female job applicants. *The Journal of psychology*, 123(1), 59-68.
- Graf, L., König, A., Enders, A., & Hungenberg, H. (2012). Debiasing competitive irrationality: How managers can be prevented from trading off absolute for relative profit. *European Management Journal*, 30(4), 386-403.
- Green, M. C., Visser, P. S., & Tetlock, P. E. (2000). Coping with accountability cross-pressures: Low-effort evasive tactics and high-effort quests for complex compromises. *Personality and Social Psychology Bulletin*, 26(11), 1380-1391.
- Griffin, D. (2013). Citizens, representatives, and the myth of the decision-making divide. *Political Behavior*, 35(2), 261-287.
- Grose, C. R., Malhotra, N., & Parks Van Houweling, R. (2015). Explaining explanations: How legislators explain their policy positions and how citizens react. *American Journal of Political Science*, 59(3), 724-743.

- Gruder, C. L. (1971). Relationships with opponent and partner in mixed-motive bargaining. *Journal of Conflict Resolution*, *15*(3), 403-416.
- Haccoun, R. R., & Klimoski, R. J. (1975). Negotiator status and accountability source: A study of negotiator behavior. *Organizational Behavior and Human Performance*, *14*(3), 342-359.
- Hagafors, R., & Brehmer, B. (1983). Does having to justify one's judgments change the nature of the judgment process?. *Organizational Behavior and Human Performance*, *31*(2), 223-232.
- Hattrup, K., & Ford, J. K. (1995). The roles of information characteristics and accountability in moderating stereotype-driven processes during social decision making. *Organizational Behavior and Human Decision Processes*, *63*(1), 73-86.
- Hayek, A. S., Toma, C., Oberlé, D., & Butera, F. (2014). The effect of grades on the preference effect: Grading reduces consideration of disconfirming evidence. *Basic and applied social psychology*, *36*(6), 544-552.
- Herbst, K. C., Leary, M. R., & McColskey-Leary, C. P. (2013). Social-evaluative influences moderate the relationship between price and perceived quality. *Social Influence*, *8*(1), 54-69.
- Hess, T. M., Germain, C. M., Swaim, E. L., & Osowski, N. L. (2009). Aging and selective engagement: The moderating impact of motivation on older adults' resource utilization. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64*(4), 447-456.
- Hess, T. M., Rosenberg, D. C., & Waters, S. J. (2001). Motivation and representational processes in adulthood: the effects of social accountability and information relevance. *Psychology and aging*, *16*(4), 629-642.
- Hirsch, B., Reichert, B. E., & Sohn, M. (2017). The impact of clawback provisions on information processing and investment behaviour. *Management Accounting Research*, *37*, 1-11.
- Hoffman, V. B., & Patton, J. M. (1997). Accountability, the dilution effect, and conservatism in auditors' fraud judgments. *Journal of Accounting Research*, *35*(2), 227-237.
- Holt, J. E., Kinchin, G., & Clarke, G. (2012). Effects of peer-assessed feedback, goal setting and a group contingency on performance and learning by 10–12-year-old academy soccer players. *Physical Education & Sport Pedagogy*, *17*(3), 231-250.
- Hoogervorst, N., De Cremer, D., & van Dijke, M. (2010). Why leaders not always disapprove of unethical follower behavior: It depends on the leader's self-interest and accountability. *Journal of Business Ethics*, *95*(1), 29-41.
- Hooper, S., & Hannafin, M. J. (1991). The effects of group composition on achievement, interaction, and learning efficiency during computer-based cooperative instruction. *Educational Technology Research and Development*, *39*(3), 27-40.

- Hoos, F., & Bollmann, G. (2012). Is accountability a double-edged sword? Experimental evidence on the effectiveness of internal controls to prevent fraud. *Journal of Management Control*, 23(2), 115-132.
- Huber, O., Bär, A. S., & Huber, O. W. (2009). Justification pressure in risky decision making: Search for risk defusing operators. *Acta Psychologica*, 130(1), 17-24.
- Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2014). Social defaults: Observed choices become choice defaults. *Journal of Consumer Research*, 41(3), 746-760.
- Huneke, M. E., Cole, C., & Levin, I. P. (2004). How varying levels of knowledge and motivation affect search and confidence during consideration and choice. *Marketing Letters*, 15(2-3), 67-79.
- Jermias, J. (2006). The influence of accountability on overconfidence and resistance to change: A research framework and experimental evidence. *Management Accounting Research*, 17(4), 370-388.
- Johnson, V. E., & Kaplan, S. E. (1991). Experimental-Evidence on the Effects of Accountability on Auditor Judgments. *Auditing-A Journal of Practice & Theory*, 10, 96-107.
- Jonas, E., Schulz-Hardt, S., & Frey, D. (2005). Giving advice or making decisions in someone else's place: The influence of impression, defense, and accuracy motivation on the search for new information. *Personality and Social Psychology Bulletin*, 31(7), 977-990.
- Joosten, A., Van Dijke, M., Van Hiel, A., & De Cremer, D. (2014). Feel good, do-good!? On consistency and compensation in moral self-regulation. *Journal of Business Ethics*, 123(1), 71-84.
- Kadous, K., & Sedor, L. M. (2004). The efficacy of third-party consultation in preventing managerial escalation of commitment: The role of mental representations. *Contemporary Accounting Research*, 21(1), 55-82.
- Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy-confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology*, 61(5), 698-707.
- Keller, J. W., & Yang, Y. E. (2008). Leadership style, decision context, and the poliheuristic theory of decision making: An experimental analysis. *Journal of Conflict Resolution*, 52(5), 687-712.
- Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, 31(2), 231-245.
- Kennedy, J. (1995). Debiasing the curse of knowledge in audit judgment. *Accounting Review*, 70(2), 249-273.

- Kim, J., Kim, J. E., & Park, J. (2012). Effects of cognitive resource availability on consumer decisions involving counterfeit products: The role of perceived justification. *Marketing Letters*, 23(3), 869-881.
- Kim, S., & Trotman, K. T. (2015). The comparative effect of process and outcome accountability in enhancing professional scepticism. *Accounting & Finance*, 55(4), 1015-1040.
- Klemz, B. R., & Gruca, T. S. (2001). Managerial assessment of potential entrants: Processes and pitfalls. *International Journal of Research in Marketing*, 18(1-2), 37-51.
- Klimoski, R. J. (1972). The effects of intragroup forces on intergroup conflict resolution. *Organizational Behavior and Human Performance*, 8(3), 363-383.
- Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes*, 45(2), 194-208.
- Klimoski, R. J., & Ash, R. A. (1974). Accountability and negotiator behavior. *Organizational Behavior and Human Performance*, 11(3), 409-425.
- Koch, C., Weber, M., & Wüstemann, J. (2012). Can auditors be independent? Experimental evidence on the effects of client type. *European Accounting Review*, 21(4), 797-823.
- Koonce, L., Anderson, U., & Marchant, G. (1995). Justification of decisions in auditing. *Journal of Accounting Research*, 33(2), 369-384.
- Krahé, B., Temkin, J., & Bieneck, S. (2007). Schema-driven information processing in judgements about rape. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(5), 601-619.
- Kramer, R. M., Pommerenke, P., & Newton, E. (1993). The social context of negotiation: Effects of social identity and interpersonal accountability on negotiator decision making. *Journal of Conflict Resolution*, 37(4), 633-654.
- Kray, L. J., Thompson, L., & Lind, E. A. (2005). It's a bet! A problem-solving approach promotes the construction of contingent agreements. *Personality and Social Psychology Bulletin*, 31(8), 1039-1051.
- Kroon, M. B., Van Kreveld, D., & Rabbie, J. M. (1992). Group versus individual decision making: Effects of accountability and gender on groupthink. *Small Group Research*, 23(4), 427-458.
- Kroon, M. B., Van Kreveld, D., & Rabbie, J. M. (1991). Police intervention in riots: The role of accountability and group norms. A field experiment. *Journal of community & applied social psychology*, 1(4), 249-267.
- Kroon, M. B., Hart, P. T., & Van Kreveld, D. (1991). Managing group decision making processes: Individual versus collective accountability and groupthink. *International Journal of Conflict Management*, 2(2), 91-115.

- Le Barbenchon, E., Milhabet, I., & Bry, C. (2016). This only happens to others: Different self-presentations in friendship and professional situations. *Social Psychology, 47*(5), 270-280.
- Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. (1999). Motivated search: Effects of choice accountability, issue involvement, and prior knowledge on information acquisition and use. *Journal of Business Research, 45*(1), 75-88.
- Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24*(6), 563-574.
- Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review, 79*(4), 1075-1094.
- Lierl, M. (2016). Social sanctions and informal accountability: Evidence from a laboratory experiment. *Journal of Theoretical Politics, 28*(1), 74-104.
- Liu, W., Friedman, R., & Hong, Y. Y. (2012). Culture and accountability in negotiation: Recognizing the importance of in-group relations. *Organizational Behavior and Human Decision Processes, 117*(1), 221-234.
- Liu, Y. C., & McLeod, P. L. (2014). Individual preference for procedural order and process accountability in group problem-solving. *Small Group Research, 45*(2), 154-175.
- Lord, A. T. (1992). Pressure: A methodological consideration for behavioral research in auditing. *Auditing, 11*(2), 90.
- Malaviya, P., & Sivakumar, K. (2002). The influence of choice justification and stimulus meaningfulness on the attraction effect. *Journal of Marketing Theory and Practice, 10*(4), 20-29.
- Markman, K. D., & Tetlock, P. E. (2000). Accountability and close-call counterfactuals: The loser who nearly won and the winner who nearly lost. *Personality and Social Psychology Bulletin, 26*(10), 1213-1224.
- Markman, K. D., & Tetlock, P. E. (2000). 'I couldn't have known': Accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology, 39*(3), 313-325.
- McAllister, D. W., Mitchell, T. R., & Beach, L. R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational behavior and human performance, 24*(2), 228-244.
- McDougall, D., & Granby, C. (1996). How expectation of questioning method affects undergraduates' preparation for class. *The Journal of experimental education, 65*(1), 43-54.

- McKenna, F. P., & Myers, L. B. (1997). Illusory self-assessments—Can they be reduced?. *British Journal of Psychology*, *88*(1), 39-51.
- Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology*, *80*(4), 517-524.
- Mero, N. P., Guidice, R. M., & Brownlee, A. L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management*, *33*(2), 223-252.
- Mitchell, T. R., Hopper, H., Daniels, D., Falvy, J. G., & Ferris, G. R. (1998). Power, accountability, and inappropriate actions. *Applied Psychology*, *47*(4), 497-517.
- Morris, M. W., & Moore, P. C. (2000). The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, *45*(4), 737-765.
- Mosterd, I., & Rutte, C. G. (2000). Effects of time pressure and accountability to constituents on negotiation. *International Journal of Conflict Management*, *11*(3), 227-247.
- Murphy, R. (1994). The effects of task characteristics on covariation assessment: The impact of accountability and judgment frame. *Organizational Behavior and Human Decision Processes*, *60*(1), 139-155.
- Nijstad, B. A., & Oltmanns, J. (2012). Motivated information processing and group decision refusal. *Group Processes & Intergroup Relations*, *15*(5), 637-651.
- Notz, W. W., & Starke, F. A. (1978). Final-offer versus conventional arbitration as means of conflict management. *Administrative Science Quarterly*, *23*(2), 189-203.
- O'Brien, B. (2009). A Recipe for Bias: An Empirical Look at the Interplay Between Institutional Incentives and Bounded Rationality in Prosecutorial Decision Making. *Missouri Law Review*, *74*(4), 999-1050.
- Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of political Economy*, *115*(2), 200-249.
- Ordóñez, L. D., Benson III, L., & Beach, L. R. (1999). Testing the compatibility test: How instructions, accountability, and anticipated regret affect prechoice screening of options. *Organizational Behavior and Human Decision Processes*, *78*(1), 63-80.
- Palmer, J. K., & Feldman, J. M. (2005). Accountability and need for cognition effects on contrast, halo, and accuracy in performance ratings. *The Journal of psychology*, *139*(2), 119-138.
- Paolini, S., Crisp, R. J., & McIntyre, K. (2009). Accountability moderates member-to-group generalization: Testing a dual process model of stereotype change. *Journal of Experimental Social Psychology*, *45*(4), 676-685.

- Patrick, S., Marsh, R., Bundy, W., Mimura, S., & Perkins, T. (2004). Control group study of juvenile diversion programs: An experiment in juvenile diversion—the comparison of three methods and a control group. *The Social Science Journal, 41*(1), 129-135.
- Pendry, L. F., & Macrae, C. N. (1996). What the disinterested perceiver overlooks: Goal-directed social categorization. *Personality and Social Psychology Bulletin, 22*(3), 249-256.
- Pennington, J., & Schlenker, B. R. (1999). Accountability for consequential decisions: Justifying ethical judgments to audiences. *Personality and Social Psychology Bulletin, 25*(9), 1067-1081.
- Peterson, E., & Thompson, L. (1997). Negotiation teamwork: The impact of information distribution and accountability on performance depends on the relationship among team members. *Organizational Behavior and Human Decision Processes, 72*(3), 364-383.
- Peytcheva, M., & Gillett, P. R. (2011). How partners' views influence auditor judgment. *Auditing: A Journal of Practice & Theory, 30*(4), 285-301.
- Pickerd, J. S., Summers, S. L., & Wood, D. A. (2014). An examination of how entry-level staff auditors respond to tone at the top vis-à-vis tone at the bottom. *Behavioral Research in Accounting, 27*(1), 79-98.
- Prentice-Dunn, S., & Rogers, R. W. (1982). Effects of public and private self-awareness on deindividuation and aggression. *Journal of Personality and Social Psychology, 43*(3), 503-513.
- Quinn, A., & Schlenker, B. R. (2002). Can accountability produce independence? Goals as determinants of the impact of accountability on conformity. *Personality and Social Psychology Bulletin, 28*(4), 472-483.
- Rausch, A., & Brauneis, A. (2015). It's about how the task is set: the inclusion–exclusion effect and accountability in preprocessing management information. *Central European Journal of Operations Research, 23*(2), 313-344.
- Rausch, A., & Brauneis, A. (2015). The effect of accountability on management accountants' selection of information. *Review of Managerial Science, 9*(3), 487-521.
- Ready, J. T., & Young, J. T. (2015). The impact of on-officer video cameras on police–citizen contacts: Findings from a controlled experiment in Mesa, AZ. *Journal of experimental criminology, 11*(3), 445-458.
- Rich, J. S. (2004). Reviewers' responses to expectations about the client and the preparer. *The Accounting Review, 79*(2), 497-517.
- Ritchie, D., & Thorkildsen, R. (1994). Effects of accountability on students' achievement in mastery learning. *The Journal of Educational Research, 88*(2), 86-90.

- Roch, S. G. (2005). An investigation of motivational factors influencing performance ratings: Rating audience and incentive. *Journal of Managerial Psychology*, 20(8), 695-711.
- Roch, S. G. (2006). Discussion and consensus in rater groups: Implications for behavioral and rating accuracy. *Human Performance*, 19(2), 91-115.
- Roggeveen, A. L., & Johar, G. V. (2004). Integration of discrepant sales forecasts: the influence of plausibility inferences based on an evoked range. *Journal of Marketing Research*, 41(1), 19-30.
- Roxβnagel, C. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, 30(3), 429-445.
- Ruscher, J. B., Wallace, D. L., Walker, K. M., & Bell, L. H. (2010). Constructive feedback in cross-race interactions. *Group Processes & Intergroup Relations*, 13(5), 603-619.
- Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. *Journal of Consulting and Clinical Psychology*, 68(1), 145-154.
- Russo, J. E., Meloy, M. G., & Wilks, T. J. (2000). Predecisional distortion of information by auditors and salespersons. *Management Science*, 46(1), 13-27.
- Schoenrade, P. A., Batson, C. D., Brandt, J. R., & Loud, R. E. (1986). Attachment, accountability, and motivation to benefit another not in distress. *Journal of Personality and Social Psychology*, 51(3), 557-563.
- Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, 43(4), 539-552.
- Schul, Y., & Mayo, R. (2003). Searching for certainty in an uncertain world: The difficulty of giving up the experiential for the rational mode of thinking. *Journal of Behavioral Decision Making*, 16(2), 93-106.
- Sedikides, C., Herbst, K. C., Hardin, D. P., & Dardis, G. J. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, 83(3), 592-605.
- Selart, M. (1996). Structure compatibility and restructuring in judgment and choice. *Organizational Behavior and Human Decision Processes*, 65(2), 106-116.
- Sengupta, J., & Johar, G. V. (2002). Effects of inconsistent attribute information on the predictive value of product attitudes: Toward a resolution of opposing perspectives. *Journal of Consumer research*, 29(1), 39-56.

- Serra, D. (2011). Combining top-down and bottom-up accountability: evidence from a bribery experiment. *The journal of law, economics, & organization*, 28(3), 569-587.
- Shankar, P. G., & Tan, H. T. (2006). Determinants of audit preparers' workpaper justifications. *The Accounting Review*, 81(2), 473-495.
- Shelton, S. W. (1999). The effect of experience on the use of irrelevant evidence in auditor judgment. *The Accounting Review*, 74(2), 217-224.
- Shore, T. H., & Tashchian, A. (2007). Effects of feedback accountability and self-rating information on employee appraisals: A replication and extension. *Psychological reports*, 100(3), 1091-1100.
- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, 65(1), 1-17.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research*, 16(2), 158-174.
- Simonson, I., & Staw, B. M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of applied psychology*, 77(4), 419-426.
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational behavior and human decision processes*, 51(3), 416-446.
- Simonson, I., Carmon, Z., & O'curry, S. (1994). Experimental evidence on the negative effect of product features and sales promotions on brand choice. *Marketing Science*, 13(1), 23-40.
- Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, 52(4), 701-717.
- Slavin, R. E., & Tanner, A. M. (1979). Effects of cooperative reward structures and individual accountability on productivity and learning. *The Journal of Educational Research*, 72(5), 294-298.
- Smith, B. (2012). *The effects of accountability on leniency reduction in self-and peer ratings on team-based performance appraisals* (Doctoral dissertation). Clemson University, Clemson, South Carolina, United States.
- Smith, J. R., Terry, D. J., & Hogg, M. A. (2007). Social identity and the attitude-behaviour relationship: Effects of anonymity and accountability. *European Journal of Social Psychology*, 37(2), 239-257.
- Staw, B. M., & Boettger, R. D. (1990). Task revision: A neglected form of work performance. *Academy of Management Journal*, 33(3), 534-559.

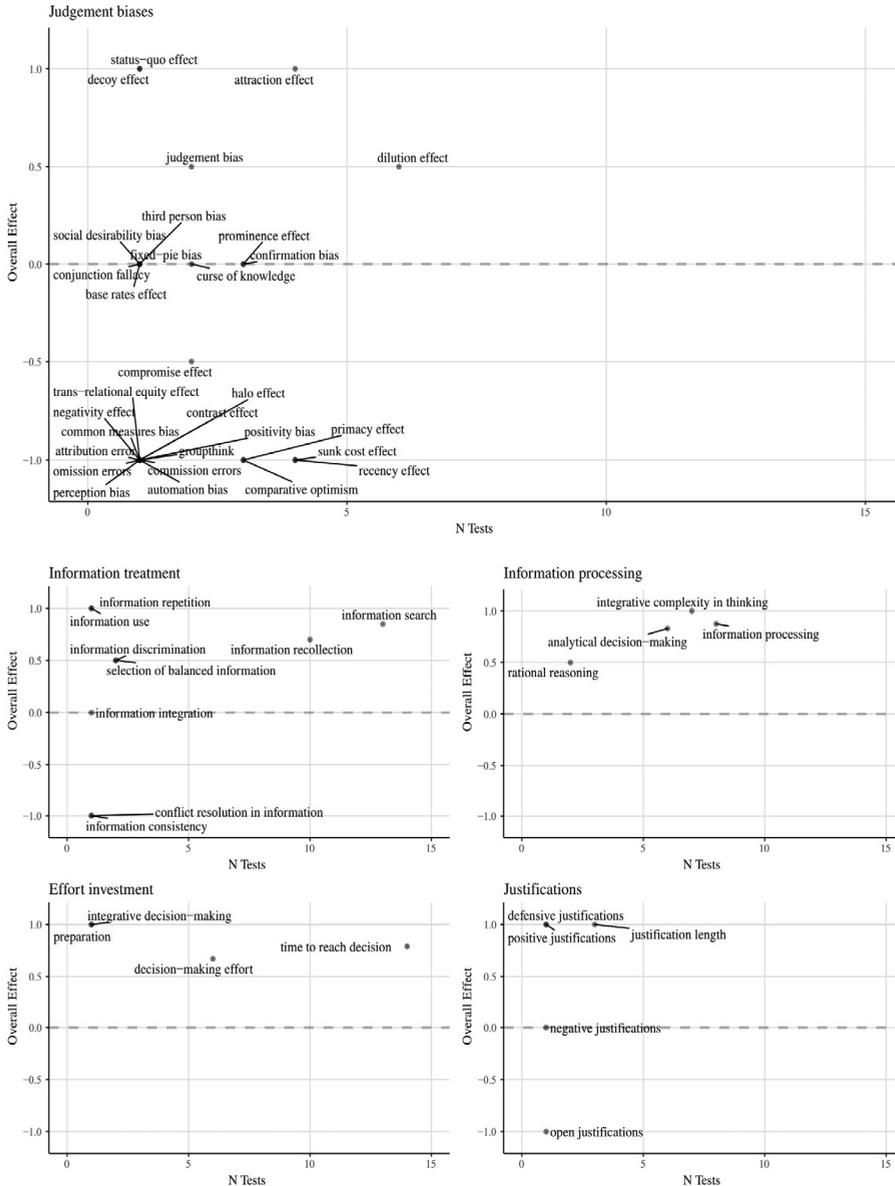
- Strohmer, D. C., & Shivvy, V. A. (1994). Bias in counselor hypothesis testing: Testing the robustness of counselor confirmatory bias. *Journal of Counseling & Development, 73*(2), 191-197.
- Swerdlin, M. R. (1992). *Motivated strategies in the performance appraisal process: Effects of rater accountability* (Doctoral dissertation). Rice University, Houston, Texas, United States.
- Tan, H. T. (1995). Effects of expectations, prior involvement, and review awareness on memory for audit evidence and judgment. *Journal of Accounting Research, 33*(1), 113-135.
- Tan, H. T., & Kao, A. (1999). Accountability effects on auditors' performance: The influence of knowledge, problem-solving ability, and task complexity. *Journal of Accounting Research, 37*(1), 209-223.
- Tan, H. T., & Jamal, K. (2006). Managing perceptions of technical competence: How well do auditors know how others view them?. *Contemporary Accounting Research, 23*(3), 761-787.
- Tan, H. T., Ng, T. B. P., & Mak, B. W. Y. (2002). The effects of task complexity on auditors' performance: The impact of accountability and knowledge. *Auditing: A Journal of Practice & Theory, 21*(2), 81-95.
- Taylor, K. A. (1995). Testing credit and blame attributions as explanation for choices under ambiguity. *Organizational Behavior and Human Decision Processes, 64*(2), 128-137.
- Ten Velden, F. S., Beersma, B., & De Dreu, C. K. (2010). It takes one to tango: The effects of dyads' epistemic motivation composition in negotiation. *Personality and Social Psychology Bulletin, 36*(11), 1454-1466.
- Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of personality and social psychology, 45*(1), 74-83.
- Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly, 46*(4), 285-292.
- Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48*(3), 227-236.
- Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of personality and social psychology, 52*(4), 700-709.
- Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of personality and social psychology, 57*(3), 388-398.
- Tetlock, P. E., & Boettger, R. (1994). Accountability amplifies the status quo effect when change creates victims. *Journal of Behavioral Decision Making, 7*(1), 1-23.

- Tetlock, P. E., Lerner, J. S., & Boettger, R. (1996). The dilution effect: judgmental bias, conversational convention, or a bit of both?. *European Journal of Social Psychology, 26*(6), 915-934.
- Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: conformity, complexity, and bolstering. *Journal of personality and social psychology, 57*(4), 632-640.
- Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology, 66*(3), 474-489.
- Thompson, L. (1995). "They saw a negotiation": Partisanship and involvement. *Journal of Personality and Social Psychology, 68*(5), 839-853.
- Turner, C. W. (2001). Accountability demands and the auditor's evidence search strategy: The influence of reviewer preferences and the nature of the response (belief vs. action). *Journal of Accounting Research, 39*(3), 683-706.
- Tyler, J. M., & Rosier, J. G. (2009). Examining self-presentation as a motivational explanation for comparative optimism. *Journal of Personality and Social Psychology, 97*(4), 716-727.
- Van der Schalk, J., Beersma, B., Van Kleef, G. A., & De Dreu, C. K. (2010). The more (complex), the better? The influence of epistemic motivation on integrative bargaining in complex negotiation. *European Journal of Social Psychology, 40*(2), 355-365.
- Van Hiel, A., & Schittekatte, M. (1998). Information Exchange in Context: Effects of Gender Composition of Group, Accountability, and Intergroup Perception on Group Decision Making. *Journal of Applied Social Psychology, 28*(22), 2049-2067.
- Van Kleef, G. A., Steinel, W., Van Knippenberg, D., Hogg, M. A., & Svensson, A. (2007). Group member prototypicality and intergroup negotiation: How one's standing in the group affects negotiation behaviour. *British Journal of Social Psychology, 46*(1), 129-152.
- Wadhwa, G., Schulz, H., & Mann, B. L. (2006). Effects of Anonymity and Accountability During Online Peer Assessment. In Mann, B. L. (Ed.), *Selected Styles in Web-Based Educational Research* (pp. 302-333). Hershey: Information Science Publishing.
- Wang, S., Noe, R. A., & Wang, Z. M. (2014). Motivating knowledge sharing in knowledge management systems: A quasi-field experiment. *Journal of Management, 40*(4), 978-1009.
- Warmke, D. L. (1979). *Effects of accountability procedures upon the utility of peer ratings of present performance* (Doctoral dissertation). The Ohio State University, Columbus, Ohio, United States.
- Webb, R. A. (2002). The impact of reputation and variance investigations on the creation of budget slack. *Accounting, Organizations and Society, 27*(4-5), 361-378.

- Webster, D. M., Richter, L., & Kruglanski, A. W. (1996). On leaping to conclusions when feeling tired: Mental fatigue effects on impressional primacy. *Journal of experimental social psychology*, 32(2), 181-195.
- Weigold, M. F., & Schlenker, B. R. (1991). Accountability and risk taking. *Personality and Social Psychology Bulletin*, 17(1), 25-29.
- Weldon, E., & Gargano, G. M. (1988). Cognitive loafing: The effects of accountability and shared responsibility on cognitive effort. *Personality and Social Psychology Bulletin*, 14(1), 159-171.
- Wells, G. L., Petty, R. E., Harkins, S. G., Kagehiro, D., & Harvey, J. H. (1977). Anticipated discussion of interpretation eliminates actor-observer differences in the attribution of causality. *Sociometry*, 40(3), 247-253.
- White, K., & Peloza, J. (2009). Self-benefit versus other-benefit marketing appeals: Their effectiveness in generating charitable support. *Journal of Marketing*, 73(4), 109-124.
- Wildschut, T., Van Horen, F., & Hart, C. (2015). Accountability accentuates interindividual-intergroup discontinuity by enforcing parochialism. *Frontiers in psychology*, 6, 1789.
- Wolff, H. G., & Moser, K. (2008). Choice, accountability, and effortful processing in escalation situations. *Zeitschrift für Psychologie/Journal of Psychology*, 216(4), 235-243.
- Yang, X. J., Wickens, C. D., Park, T., Fong, L., & Siah, K. T. (2015). Effects of information access cost and accountability on medical residents' information retrieval strategy and performance during prehandover preparation: evidence from interview and simulation study. *Human factors*, 57(8), 1459-1471.
- Yarnold, P. R., Mueser, K. T., & Lyons, J. S. (1988). Type A behavior, accountability, and work rate in small groups. *Journal of Research in Personality*, 22(3), 353-360.
- Yun, G. J., Donahue, L. M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment*, 13(2), 97-107.

2. C. Detailed representation of the recorded effects

Figure 2.C.1 *Effects of accountability on five aspects of decision-making*



Note: The x-axis represents the number of times the effect has been tested experimentally, while the y-axis presents the direction of the effect (positive effects appear on the upper part of the graph, while negative at the bottom part of the graph), as well as the robustness of the effects (the



placement of the variable on the y-axis is a proportion of the sum of the effects across studies with 1 being positive, -1 negative, and 0 no effect, and the number of studies, for example, if an effect has been tested three times and twice found positive, and once no effect was found, the effect is calculated as $(1+1+0)/3=0.67$.

Figure 2.C.2 *Effects of accountability pressure on six aspects of behaviour*

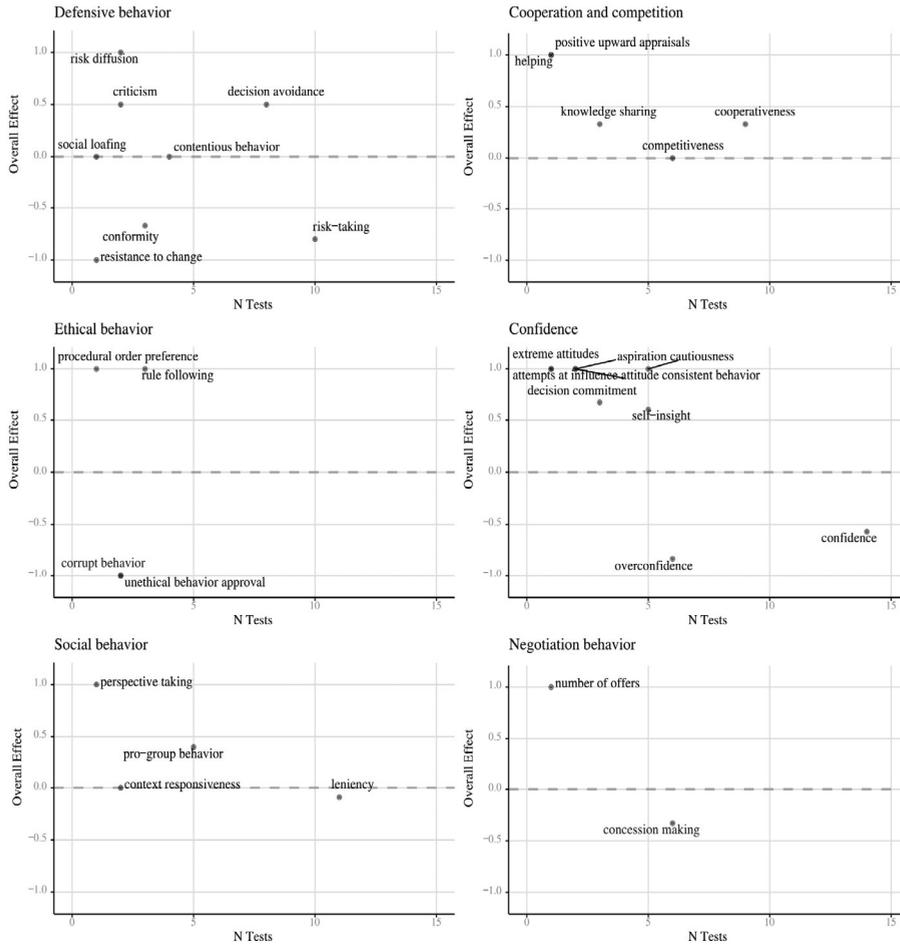
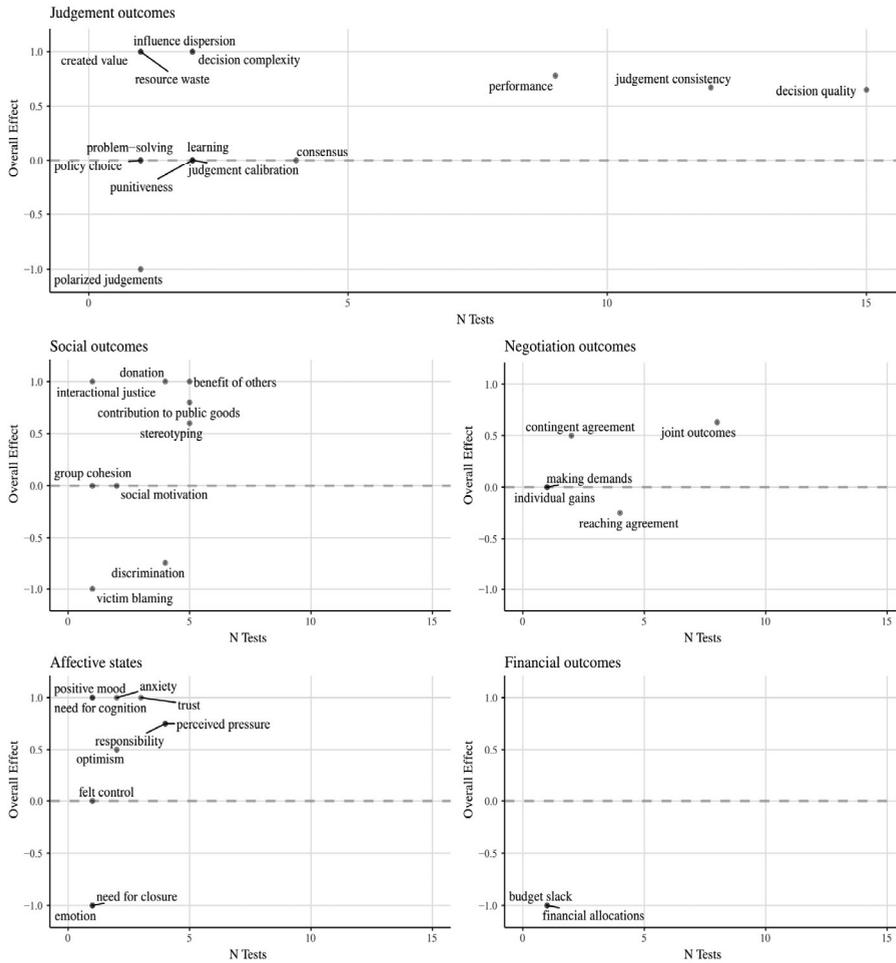


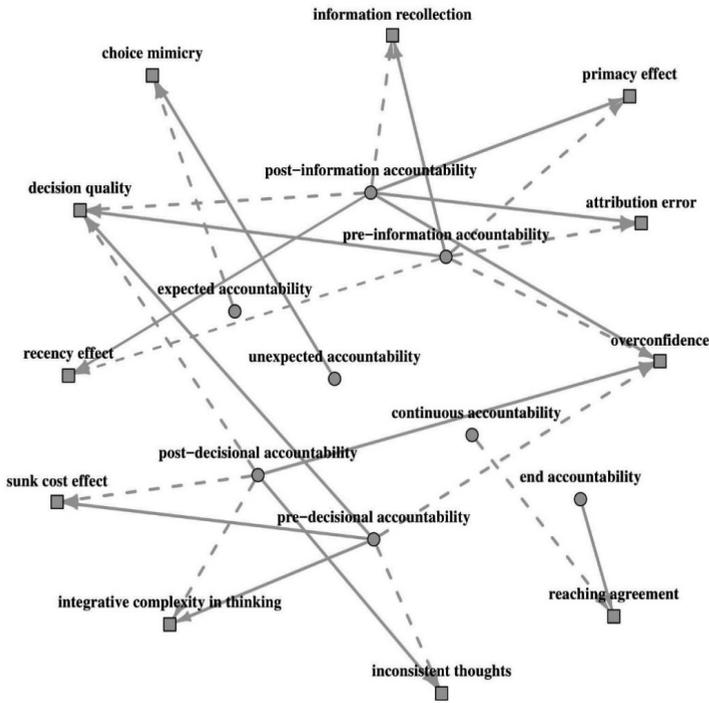
Figure 2.C.3 Effects of accountability on five types of outcomes



Note: Decision quality has been actually experimentally tested 26 times, but for ease of representation it has been placed at the maximum on the displayed scale.



Figure 2.C.4 *Effects of the characteristics of accountability mechanisms on various tested outcomes*



a) Effects of accountability timing



b) Effects of accountability standard



2.D. Meta-analysis on the main effects of accountability on decision-making

The systematic analysis of the behavioral and experimental literature on accountability identified several well-tested effects of accountability on decision-making behavior. As most of them appear to point in the same direction, here we estimate the size of their effects using a meta-analysis. The six effects which are subject to this meta analysis are *decision-making time*, *information search*, *information recollection*, *integrative complexity in thinking*, *decision consistency*, and *decision accuracy*. As each of these dimensions could be seen as normatively desirable aspects of decision-making, they are jointly considered as dimensions of decision-making quality.

Method. The method of meta-analysis is a useful tool for summarizing the findings of a larger literature empirically investigating a particular question or relationship. The meta-analysis provides more precise estimate of a particular effect, by calculating a weighted average of the effects of multiple studies. It also presents a useful tool for detecting heterogeneity in the effect sizes across samples, uncovering subpopulations and moderators affecting the size of the effect (Card, 2015; Field, 2001). Here, meta-analysis is employed with the goal of estimating the size of the effect of accountability pressure on six aspects of the decision-making process. The calculation of the meta-analysis estimate was performed using the tool Meta Essentials (Suurmond et al., 2017). Meta Essentials is a simple validated software which allows for the calculation of effect sizes from different statistics. It also offers the opportunity to perform subgroup, moderator, and publication bias analyses (Suurmond et al., 2017).

Given that the corpus of studies includes only experiments and the effects of interest primarily represent differences between means of experimental groups, the measure of Hedge's g was used to represent the effect sizes in the selected studies. The measure of Hedge's g provides a correction to Cohen's d bias in small sample sizes, and thus it is generally preferred (Lakens, 2013). The

calculation of the effects was performed with the help of the calculation sheet provided by Lakens (2013), using the statistical information provided in the individual studies, such as F and t test results, sample sizes, and means and standard deviations. When discussing the effect sizes, we rely on the classical Cohen's classification, considering $d = 0.2$ small, $d = 0.5$ medium and $d = 0.8$ a large effect size (Cohen, 1988).

Some studies consist of multiple experiments with differences in their designs and samples. If more than one experiment in a single study studied the same relationship of interest, using however different experimental designs and samples, the effects were recorded separately. However, when multiple indices of the same effect, or multiple similar effects were reported resulting from the same experiment, the simple average effect size was calculated and used, due to the dependency of the effects (Borenstein et al., 2009; Moeyaert et al., 2017). Thus, when the overall effect of accountability on decision-making quality was estimated, the effect sizes of the various sub-effects found in a same experiment were averaged.

Few of the studies did not provide sufficient information to estimate the effect size. The authors of those studies were contacted with request for additional information²⁸, however, these requests did not result in additional information. One author responded that no additional information was available beyond what was reported in the study, one author did not respond, and one author responded that his study might not be fit for our purposes (with which we later agreed). In some studies the information was incomplete to precisely estimate the effect size. For instance, in some experiments the exact number of participants per group was not reported. In those cases, the unbiased estimate Hedge's g was calculated by imputing an equal split of the total number of

²⁸ Only the authors of one study were not contacted due to the fact that the study was published more than 25 years ago.

experimental participants per group. Lastly, in several cases the effect was reported as not statistically significant, but no sufficient information was provided to estimate the exact effect size.

In those cases, the effect was imputed as zero.

The meta-analysis estimation can be performed with either a fixed-effect model or a random-effects model. Using the fixed-effect model to estimate the meta-effect implies that we assume that all studies in the analysis share a common effect size. The differences in effect sizes between individual studies are considered to be due to sampling error. The meta-analytic estimate in a fixed-effect model thus presents the common effect size for all studies. Conversely, using the random-effects model implies that the effect sizes differ among studies, and there is thus a distribution of true effect sizes. These differences in effect sizes could arise due to differences in the study participants, and differences in study interventions, among other things. Thus, the meta-analytic estimate here aims to determine the mean effect size of the studies (Borenstein et al., 2009; Borenstein et al., 2010). In this study, the analysis was performed using a random-effects model. Since the studies included in this meta-analysis have been performed on different types of participants, in different types of experimental settings, and employing somewhat different experimental manipulations, there is little to no basis to assume that they will have a common effect size (Borenstein et al., 2009; Borenstein et al., 2010; Field & Gillett, 2010).

Results. Does accountability improve decision-making quality? Numerous studies suggest that they do (Lerner & Tetlock, 1999), however, whether the observed improvement is consistent and substantive, it is not yet clear. This meta-analysis offers insights about the magnitude of the effects of accountability on quality on decision-making. Table 2.D.1 presents the overview of the meta-effects of accountability pressures on decision-making quality.

The overall analysis shows that the pressure to justify one’s decision does have a substantive and significant effect on decision-making quality. The overall effect is Hedge’s $g = 0.48$ which points out to a medium-sized effect. The sub meta-analyses show the effects of accountability per dimension of decision-making quality. The results indicate that all dimensions are positively and significantly affected by accountability pressures, however, the effects are stronger for decision-making time, information search, integrative complexity in thinking and decision consistency (Hedge’s $g > 0.5$), than for decision recollection and decision accuracy (Hedge’s $g > 0.2$ and < 0.4).

Table 2.D.1 *Overview of results*

Analysis	Hedge’s g	CI _L	CI _u	PI _L	PI _u	I ²	N	K
<i>Overall analysis</i>								
Decision quality	0.48	0.38	0.58	-0.02	0.98	57.74%	4553	46
<i>Sub meta-analysis</i>								
Decision-making time	0.57	0.44	0.70	0.41	0.72	2.95%	1214	14
Information search	0.53	0.26	0.80	-0.26	1.32	71.11%	1095	12
Information recollection	0.45	-0.05	0.95	-0.99	1.89	85.47%	624	7
Integrative complexity	0.52	0.33	0.70	0.23	0.81	22.54%	1071	7
Decision consistency	0.53	0.36	0.70	0.36	0.70	0.00%	699	8
Decision accuracy	0.23	0.04	0.42	-0.28	0.73	51.64%	1150	11

Moderator analysis. The main analysis indicates that accountability as a mechanism is an important tool for influencing human behavior – it increases decision making effort and overall quality. However, the main analysis also indicates that substantive heterogeneity can be found in both the overall effect and the majority of the sub-effects, as displayed by I², which implies that the observed effects could be influenced by some of the study characteristics. Thus, moderator

analysis is an appropriate tool to investigate the sources of this observed heterogeneity (Borenstein et al., 2009).

More substantively, accountability is a pervasive mechanism found in many social, organizational and political contexts. While in social contexts it appears as a regulator of human relations, and often an indicator of social influence, in organizational and political context it is purposefully set up to assure that the behavior of certain individuals and groups complies to certain expected norms. In public organizations and political posts, this is of even greater importance, since the individuals and groups subjected to the accountability mechanisms are entrusted powers with the goal of serving the public (Bovens, 2010). Thus, it is of paramount importance to discern the different effects accountability mechanisms have in different contexts and among different types of individuals, in order to understand and better anticipate its effects (Aleksavska et al., 2019). Furthermore, an analysis of the characteristics of existing accountability studies and its effects could point out to important gaps in the literature. Thus, if the majority of the studies are conducted in only one type of experimental setting, or on one type of participants, it might be fruitful to replicate these studies on other participants in other settings in order to obtain a better grasp of the different effects of the phenomenon.

The experimental studies on accountability included in this meta-analysis are performed in a variety of settings, using different experimental tools and participants (Aleksavska et al., 2019). Thus, they differ in the type of experiment, the type of participants, as well as the type of decision-making task. The way these differences affect the results from these experimental studies is investigated in a moderator analysis. The results of the analysis are presented in Table 2.D.2 below.

Researchers have used different types of experimental settings to investigate the effects of accountability on decision-making quality. However, laboratory experiments seem to be the

dominant choice. Controlled experiments performed outside of the laboratory or classroom, or lab-in-the-field experiments, as well as field experiments, online and mailed experiments are only sparsely used. The moderator analysis shows that lab-in-the-field and field experiments detect marginally stronger effects of accountability on decision-making quality than laboratory experiments, while online and mail experiments detect considerably weaker effects than both laboratory and lab-in-the-field experiments. Thus, the experimental setting seems to have an influence on the observed effects.

The experiments were performed on different types of experimental participants as well, however, not surprisingly, the majority have been performed on students. A small portion of the experiments have been performed on professionals from different sectors, citizens or consumers, or a mix²⁹. The moderator analysis indicates that the effects of accountability on decision-making quality is marginally stronger among students than among professionals, and considerably stronger than among citizens or consumers. Thus, the type of experimental participants seems to also have an effect on the size of the accountability effect on decision-making quality.

While most accountability experiments are performed in an individual setting – one person performing the task individually – a number of experiments are conducted in small groups of several participants, usually two or three. Thus, in these group experiments, the accountability pressure is dispersed on several individuals, as opposed to only one. It is likely that this difference in accountability pressure will affect the effect of accountability on decision-making quality. The moderator analysis indicates that the effect of accountability pressure is somewhat stronger if directed towards a single individual than towards a group. Thus, whether the bearer of

²⁹ One study did not report the type of experimental participants (Schul & Mayo, 2003).

accountability pressure is a single person or a group does seem to affect the size of the effect accountability has on decision-making quality.

Table 2.D.2 Moderator analysis

Moderator	Hedge 's g	CI _L	CI _u	PI _L	PI _u	I ²	N	K
Experimental setting								
Laboratory	0.50	0.39	0.61	-0.04	1.04	60.89%	3815	39
Lab in the field / field	0.53	0.30	0.77	0.15	0.91	0.00%	338	4
Online/mail	0.39	0.26	0.52	-0.43	1.21	0.00%	400	2
Participants								
Students	0.50	0.39	0.62	-0.05	1.06	62.68%	3375	34
Professionals	0.47	0.29	0.64	0.24	0.70	0.00%	488	6
Citizens/consum ers	0.60	0.16	1.04	-0.94	2.14	58.70%	228	3
Mixed	0.23	-0.19	0.66	-4.13	4.59	76.51%	422	2
Type of decision-making								
Individual	0.49	0.38	0.60	-0.04	1.02	62.15%	4157	39
Group	0.45	0.24	0.65	0.91	0.70	0.92%	396	7

In sum, the characteristics of the experimental study in which accountability has been investigated do affect its observed effects. However, the studies included in this meta-analysis seem to be quite similar, and thus predominantly laboratory studies, with individual decision-making tasks, performed on students. Therefore, they present only limited opportunities to study the true differences in effects caused by experimental characteristics. Specifically, many more experiments on professionals, both from the public sector and the private sector are necessary to determine the effect of accountability on their decision-making processes and outcomes. Furthermore, more studies outside the laboratory or the classroom are needed, preferably capturing much more realistic and consequential situations of decision-making under accountability pressures.

Publication bias. The well-known file drawer problem, or the tendency to publish only statistically significant results, poses a threat to the validity of meta-analytic results (Rosenthal, 1979). This publication bias could lead to overestimation of the observed effects, since null results are less likely to be published, and thus also less likely to be included in a meta-analytic analyses (Borenstein et al., 2009). Thus, analyzing and discussing the potential presence of publication bias in a meta-analysis is an important step to assess the validity of its results and conclusions.

The studies included in this meta-analysis have been based on a comprehensive literature review, which also includes unpublished studies, specifically, dissertations and theses (Borenstein et al., 2009). However, this is not a guarantee that all relevant unpublished studies have been captured, since they would still have to appear in one of the five scientific databases which have been searched during the systematic literature search. One facilitating factor in this meta-analysis is that the analyzed experiments usually report several examined effects from a single study, among which null effects appear. Null effects, thus, are less likely to be “hidden” and unpublished if they appear among some significant effects, which is something we can observe in the studies in this meta-analysis. In sum, while there are some factors that reduce the threat of publication bias in this study, its presence cannot be completely ruled out.

Several methods and techniques can aid the detection of publication bias in a meta-analysis. Among them, the funnel plot has been most commonly used (Light & Pillemer, 1984). The funnel plot displays the effect sizes against their standard error (or another measure of precision). In an unbiased sample, the study effects will be distributed symmetrically around the overall meta-analytic effect in a shape of a funnel (Field & Gillet, 2010). While the funnel plot can help detect the presence of a publication bias, it does not offer tools to correct for it. The trim and fill method trims the extreme positive small size studies, estimates the effects of the missing studies on the

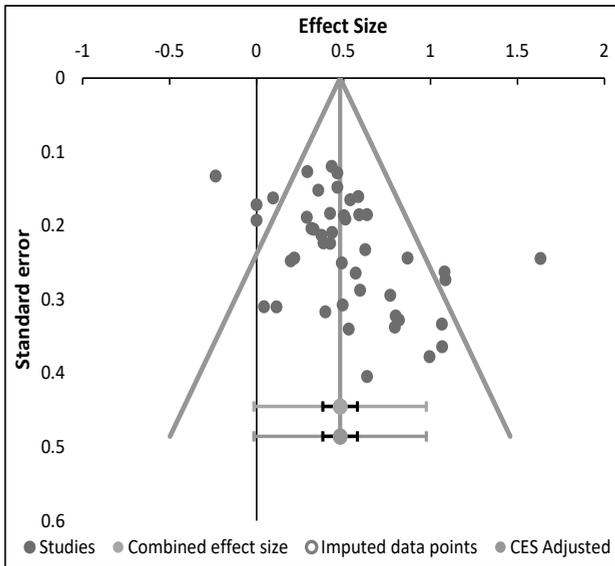
basis of the funnel plot, and provides a corrected estimate of the overall effect (Borenstein et al., 2009, Duval & Tweedie, 2000; Field & Gillet, 2010).

The results of the funnel plot and trim and fill method are presented below in Figure 2.D.1 and Table 2.D.3. The funnel plot, and the trim and fill method indicate no presence of publication bias on the overall effect of decision-making quality. They do, however, indicate possible publication bias in the studies of the sub effects of information search, integrative complexity in thinking and decision consistency. The trim and fill method imputes one data point for the sub effect of information search, and estimates the effect at Hedge's g 0.26, while for the sub effects of integrative complexity in thinking and decision consistency imputes two data points, estimating the effects at Hedge's g 0.47 and 0.41 accordingly. Thus, the results do indicate some presence of publication bias, however, only detected at the sub-effect level. The adjusted effect sizes resulting from the trim and fill method do not differ dramatically from the ones originally estimated, except for the sub effect of information search.

Table 2.D.3 *Trim and fill method analysis*

Analysis	Trim and fill method	
	Imputed data points	Adjusted effect size (Hedge's g)
<i>Overall analysis</i>		
Decision quality	0	0.48
<i>Sub meta-analysis</i>		
Decision-making time	0	0.57
Information search	1	0.26
Information recollection	0	0.45
Integrative complexity	2	0.47
Decision consistency	2	0.41
Decision accuracy	0	0.23

Figure 2.D.1 *Funnel plot on overall effect of decision quality*



Conclusion. This study presents a meta-analysis on the effects of accountability pressures on six dimensions of decision-making quality. If we rely on Cohen’s original effect size interpretation, the results of the analysis indicate that for all six investigated variables, except decision accuracy, accountability has a moderate positive effect (Cohen, 1988). This is a strong indication that accountability pressures can make important positive differences in human decision processes.

There are a few limitations to this study that should be mentioned. First, the number of included effects, as well as the sample sizes of the individual studies are not particularly large. This has some relevant consequences. In particular, the small samples have been found to display more extreme effects, which could introduce a certain bias in the results. This limitation could be overcome only with significantly larger-scale replication efforts. Furthermore, the small simple sizes could also produce bias in the heterogeneity statistic. The heterogeneity statistic I^2 has been

found to have significant positive bias in small samples, which can lead to great overestimation of the true heterogeneity (von Hippel, 2015).

Table 2.D.4 *Studies included in the meta-analysis*

#	Study name	Hedges' g	CI Lower limit	CI Upper limit	Weight
	Decision-making time				
1	Rausch, A., Brauneis, A. 2015	0.54	0.25	0.83	16.59%
2	Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. 1999	0.54	0.02	1.07	5.35%
3	Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. 1999	0.69	0.17	1.22	5.41%
4	Van Hiel A., Schittekatte M. 1998	0.96	0.13	1.86	2.14%
5	Siegel-Jacobs, K., & Yates, J. F. 1996	0.59	0.15	1.05	7.26%
6	Swerdlin, M. R. 1992	0.63	0.17	1.10	6.87%
7	Kassin, S. M., Rigby, S., & Castillo, S. R. 1991	1.08	0.57	1.62	5.40%
8	Ford, J. K., & Weldon, E. 1981	0.77	0.19	1.37	4.30%
9	McAllister, D. W., Mitchell, T. R., & Beach, L. R. 1979	0.50	0.14	0.88	10.62%
10	Breaugh, J. A., & Klimoski, R. J. 1977	0.12	-0.51	0.75	3.88%
11	Bartunek, J. M., Benton, A. A., & Keys, C. B. 1975	0.87	0.39	1.36	6.25%
12	Klimoski, R. J., & Ash, R. A. 1974	0.20	-0.30	0.69	6.06%
13	Klimoski, R. J. 1972	0.32	-0.09	0.72	8.93%
14	DeZoort, T., Harrison, P., & Taylor, M. 2006	0.51	0.15	0.87	10.93%
	Information search				
1	Rausch, A., Brauneis, A. 2015	0.39	0.10	0.68	10.42%
2	Griffin, D. 2013	0.00	-0.34	0.34	9.86%
3	Ten Velden, F. S., Beersma, B., & De Dreu, C. K. 2010	0.50	-0.12	1.13	6.78%
4	Huber, O., Bär, A. S., & Huber, O. W. 2009	1.09	0.55	1.65	7.49%
5	Huber, O., Bär, A. S., & Huber, O. W. 2009	0.42	-0.02	0.87	8.62%
6	Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. 2007	0.67	0.10	1.26	7.16%
7	Huneke, M. E., Cole, C., & Levin, I. P. 2004	0.37	-0.05	0.80	8.87%
8	Klemz, B. R., & Gruca, T. S. 2001	0.29	-0.08	0.66	9.45%
9	Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. 1999	0.60	0.08	1.14	7.68%
10	Van Hiel A., Schittekatte M. 1998	0.31	-0.50	1.14	5.22%
11	Webster, D. M., Richter, L., & Kruglanski, A. W. 1996	1.63	1.16	2.13	8.14%
12	Hattrup K., Ford J.K. 1995	0.36	0.06	0.66	10.31%
	Information recollection				
1	Yang, X. J., Wickens, C. D., Park, T., Fong, L., & Siah, K. T. 2015	1.00	0.25	1.80	11.98%
2	De Dreu, C. K., Beersma, B., Stroebe, K., & Euwema, M. C. 2006	0.82	0.17	1.50	13.00%
3	Roch, S. G. 2006	0.38	0.06	0.71	16.18%
4	McDougall, D., & Granby, C. 1996	1.07	0.41	1.76	12.89%

5	Thompson, L. 1995	-0.46	-0.73	-0.20	16.63%
6	Swerdlin, M. R. 1992	0.22	-0.27	0.71	14.72%
7	Tetlock, P. E. 1983b	0.49	0.00	1.00	14.59%
	Integrative complexity in thinking				
1	Green, M. C., Visser, P. S., & Tetlock, P. E. 2000	0.54	0.22	0.87	15.43%
2	Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. 1999	0.67	0.30	1.04	12.74%
3	Tetlock P.E., Lerner J.S., Boettger R. 1996	0.29	0.04	0.54	22.43%
4	Girvan, E. J., Deason, G., & Borgida, E. 2015	0.43	0.20	0.67	24.13%
5	Tetlock, P. E., & Boettger, R. 1989	0.59	0.27	0.91	16.06%
6	Tetlock P.E., Kim J.I. 1987	0.80	0.17	1.47	4.90%
7	Tetlock, P. E. 1983a	1.07	0.39	1.78	4.32%
	Decision consistency				
1	Arkes, H. R., González-Vallejo, C., Bonham, A. J., Kung, Y. H., & Bailey, N. 2010	0.43	0.02	0.85	14.49%
2	DeZoort, T., Harrison, P., & Taylor, M. 2006	0.76	0.39	1.13	18.23%
3	Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. 1999	0.39	0.03	0.75	18.92%
4	Ordóñez, L. D., Benson III, L., & Beach, L. R. 1999	0.51	0.14	0.89	17.36%
5	Siegel-Jacobs, K., & Yates, J. F. 1996	1.07	0.35	1.84	4.78%
6	Ashton, R. H. 1992	0.70	0.02	1.42	5.35%
7	Johnson, V. E., & Kaplan, S. E. 1991	0.33	-0.08	0.74	15.11%
8	Tetlock, P. E. 1983a	0.52	-0.14	1.19	5.75%
	Decision accuracy				
1	Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. 2007	0.52	-0.05	1.10	6.73%
2	Roch, S. G. 2006	-0.19	-0.51	0.13	12.03%
3	Palmer, J. K., & Feldman, J. M. 2005	0.42	0.06	0.79	10.92%
4	Schul, Y., & Mayo, R. 2003	0.04	-0.58	0.67	6.04%
5	Ruscio J. 2000	0.00	-0.38	0.38	10.44%
6	Siegel-Jacobs, K., & Yates, J. F. 1996	0.18	-0.26	0.63	9.09%
7	Thompson, L. 1995	-0.01	-0.27	0.25	13.80%
8	Mero, N. P., & Motowidlo, S. J. 1995	0.47	0.21	0.72	14.03%
9	Ashton, R. H. 1992	0.36	-0.32	1.05	5.36%
10	Tetlock P.E., Kim J.I. 1987	0.80	0.16	1.47	5.71%
11	Warmke, D. L. 1979	0.40	-0.23	1.04	5.86%

Appendix Chapter 3

3.A. Experimental scenario in Dutch

Wij vragen u om uzelf in te leven in de onderstaande denkbeeldige situatie:

De USBO en de stad Utrecht zijn een samenwerkingsverband aangegaan. Het hoofddoel van het project is om plastic afval in de stad te verminderen.

Als onderdeel van de samenwerking is aan masterstudenten van de USBO gevraagd om met ideeën te komen die hieraan kunnen bijdragen. De stad Utrecht is van plan om het beste projectidee verder uit te werken en vervolgens uit te voeren.

Er zijn drie voorstellen naar voren gekomen en u bent ingeschakeld als expert om uw oordeel te geven over het beste projectidee.

Welke project kiest u?

No rules	General rules	Specific rules
<p>Let op: u wordt gevraagd om uw persoonlijke voorkeur opzij te zetten bij uw keuze.</p> <p>Utrecht en de USBO vragen u als USBO-alumnus om op grond van uw professionele expertise het beste projectidee te kiezen.</p> <p>U wordt hierna gevraagd om een keuze te maken voor een project en om deze keuze uit te leggen.</p>	<p>Let op: u wordt gevraagd om uw persoonlijke voorkeur opzij te zetten bij uw keuze.</p> <p>Utrecht en de USBO hebben drie criteria opgesteld waaraan het beste projectidee moet voldoen.</p> <p>U wordt expliciet gevraagd om de onderstaande criteria te hanteren bij uw keuze voor het beste projectidee:</p> <ol style="list-style-type: none"> 1. Deelname van diverse stakeholders 2. Efficiëntie 3. Duurzaamheid van projectopbrengsten <p>U wordt hierna gevraagd om een keuze te maken voor een project en om deze keuze uit leggen op basis van deze criteria.</p>	<p>Let op: u wordt gevraagd om uw persoonlijke voorkeur opzij te zetten bij uw keuze.</p> <p>Utrecht en de USBO hebben drie criteria opgesteld waaraan het beste projectidee moet voldoen.</p> <p>U wordt expliciet gevraagd om de onderstaande criteria te hanteren bij uw keuze voor het beste projectidee:</p> <ol style="list-style-type: none"> 1. Directe deelname van zowel burgers als ondernemers. De deelname hoeft niet per se gelijk te zijn. 2. De hoogste verwachte impact op plasticreductie per uitgegeven euro. 3. De opbrengsten van het project zijn blijvend, ook

		<p>nadat het project is afgerond.</p> <p>U wordt hierna gevraagd om een keuze te maken voor een project en om deze keuze uit leggen op basis van deze criteria.</p>
--	--	--

Voorstel: Verminderen plastic verpakkingen in winkels

Wij willen winkels aanmoedigen om minder plastic verpakkingsmateriaal te gebruiken door winkel te laten deelnemen aan een zes maanden lange publieke campagne. Aan het einde worden drie winnaars publiekelijk bekend gemaakt:

- "De Groenste": grootste percentage producten zonder plastic verpakking
- "De Vergroener": grootste reductie van plastic verpakking
- "Groene innovator": beste alternatief voor plastic verpakking

Samenvatting:

Output: Campagne van zes maanden

Doelgroep: burgers en lokale ondernemers

Impact: 30.000 kg minder plastic afval in 2018

Kosten: 120.000 Euro

Voorstel: Aanmoedigen van "plasticvrije levensstijl"

De meeste producten uit supermarkten zijn verpakt in plastic dat niet geschikt is voor recycling. We willen mensen bewust maken van dit probleem door:

- een "plastic challenge" te starten: mensen uitdagen om 1 dag, week of maand geen producten met plastic verpakking te gebruiken
- een online platform te ontwikkelen met tips hierover en een database met ondernemers die het initiatief ondersteunen.

Samenvatting:

Output: campagne van zes maanden, permanente website met informatie over burgers en ondernemers

Doelgroep: burgers en lokale ondernemers

Impact: 10.000 kg minder plastic afval in 2018

Kosten: 60.000 Euro

Voorstel: Makkelijke toegang tot drinkwater "on the go"

De gemiddelde Nederlander koopt 80 plastic flessen per jaar, waarvan de helft flesjes water. We willen dit aantal terugbrengen door:

- Aanleg van fonteinen met drinkwater door de hele stad
- Campagne voeren om lokale ondernemers aan te moedigen om "hervulpunten" aan te bieden aan winkelend publiek en om gratis kraanwater aan te bieden in restaurants. Deelnemende ondernemers krijgen een zichtbare sticker en worden gepromoot op een online kaart op een website.

Samenvatting:

Output: campagne van drie maanden, waterfontijnen in elke buurt, online kaart

Doelgroep: burgers en lokale ondernemers

Impact: 20.000 kg minder plastic in 2018

Kosten: 80.000 Euro

3.B. Experimental scenario translated to English

We ask you to place yourself in the fictional situation described below:

USG and the city of Utrecht have entered into a partnership. The main goal of the project is to reduce plastic waste in the city.

As part of the collaboration, USG master students were asked to come up with ideas that could contribute to this. The city of Utrecht plans to further develop and then implement the best project idea.

Three proposals have emerged and you have been called in as an expert to give your opinion on the best project idea.

Which project do you choose?

No rules	General rules	Specific rules
<p>Please note you will be asked to put aside your personal preferences in your choice.</p> <p>Utrecht and USG ask you as USG alumni to choose the best project idea based on your professional expertise. You will then be asked to make a choice for a project and to explain this choice.</p>	<p>Please note you will be asked to put aside your personal preferences in your choice.</p> <p>Utrecht and USG have drawn up three criteria that the best project idea must meet.</p> <p>You are explicitly asked to use the criteria below when choosing the best project idea:</p> <p>1. Participation of various stakeholders 2. Efficiency 3. Durability of project returns</p> <p>You will then be asked to make a choice for a project and to explain this choice.</p>	<p>Please note you will be asked to put aside your personal preferences in your choice.</p> <p>Utrecht and USG have drawn up three criteria that the best project idea must meet.</p> <p>You are explicitly asked to use the criteria below when choosing the best project idea:</p> <p>1. Direct participation of both citizens and entrepreneurs. Participation does not necessarily have to be the same. 2. The highest expected impact on plastic reduction per euro spent. 3. The returns of the project are permanent, even after the project has been completed.</p> <p>You will then be asked to make a choice for a project and to explain this choice.</p>

Proposal: Reduce plastic packaging in shops

We want to encourage retailers to use less plastic packaging by inviting them to participate in a six-month long public campaign. At the end of the campaign three winners will be made publicly known:

- "The Greenest": largest percentage of product without plastic packaging
- "The Greener": largest reduction in plastic packaging
- "Green innovator": best alternative for plastic packaging

Summary:

Output: six-month long campaign

Target group: citizens and local entrepreneurs

Impact: 30.000 kg less plastic waste in 2018

Cost: 120.000 Euro

Proposal: Encouraging "plastic-free lifestyle"

Most products in supermarkets are packaged in plastic that is not suitable for recycling. We want to make people aware of this problem by:

- starting a "plastic challenge": challenge people not to use products with plastic packaging for 1 day, week or month;
- developing an online platform with tips about having a "plastic-free lifestyle" and a database with entrepreneurs who support this initiative.

Summary:

Output: six-month long campaign, permanent website with information for information for citizens and entrepreneurs

Target group: citizens and local entrepreneurs

Impact: 10.000 kg less plastic waste in 2018

Cost: 60.000 Euro

Proposal: Easy access to drinking water "on the go"

The average Dutch person buys 80 plastic bottles a year, half of which are bottled water. We want to reduce that number by:

- Construction of drinking water fountains throughout the city;

- Campaign to encourage local entrepreneurs to offer "refill points" to shoppers and to offer free tap water in restaurants. Participating entrepreneurs receive a visible sticker and are promoted on an online map on a website.

Summary:

Output: three-month long campaign, water fountains in every neighborhood, online map

Target group: citizens and local entrepreneurs

Impact: 20.000 kg less plastic waste in 2018

Cost: 80.000 Euro

3.C. Characteristics of experimental subjects (Study 1)

Table 3.C.1 *Overview of experimental panel characteristics*

	Frequency	Percentage
Gender		
<i>Male</i>	269	42,5%
<i>Female</i>	364	57,5%
Age group		
<i>20 to 29</i>	218	34,5%
<i>30 to 39</i>	254	40,2%
<i>40 to 49</i>	39	6,2%
<i>50 to 59</i>	80	12,6%
<i>60 and older</i>	41	6,5%
Type of work public or not		
<i>Public sector</i>	358	56,6%
<i>Other</i>	275	43,4%
Total	633	

Table 3.C.2 Differences between treatment and control groups on basic characteristics

	Full sample	No rules (control group)	General rules treatment	Specific rules treatment	Test of difference between groups
Type of work					$X^2 (12, N = 151) = 11.39, p = 0.49$
<i>Government</i>	41	18	8	15	
<i>Other public sector</i>	53	16	20	17	
<i>Semi-public organization</i>	8	4	0	4	
<i>Private company</i>	33	14	12	7	
<i>Self-employed</i>	5	1	2	2	
<i>Not working</i>	8	2	4	2	
<i>Other</i>	3	1	1	1	
Type work public or other					$X^2 (2, N = 151) = 2.58, p = 0.28$
<i>Public sector</i>	102	38	28	36	
<i>Other</i>	49	18	19	12	
Plastic issue importance					$F (2; 148) = 0.61, p = 0.55$
<i>Mean</i>	1.61	1.71	1.51	1.58	
<i>SD</i>	0.95	1.14	0.91	0.74	
N	151	56	47	48	

Note. The perception of importance of the issue of plastic was measured on a 7-point scale ranging from (1) “very important” to (7) “not important at all”.

3.D. Characteristics of experimental subjects (Study 2)

Table 3.D.1 Differences between treatment and control groups on basic characteristics

	Full sample	No rules (control group)	Specific rules treatment	Test of difference between groups
Gender				$X^2 (2, N = 84) =$
<i>Male</i>	28	16	12	2.44, $p = 0.29$
<i>Female</i>	51	28	23	
<i>NA</i>	5	1	4	
Experimental session				$X^2 (4, N = 84) =$
1	17	9	8	0.18, $p = 0.99$
2	13	7	6	
3	22	12	10	
4	18	9	9	
5	14	8	6	
Age				$F (1; 76) = 1.65, p = 0.20$
<i>Mean</i>	22.13	22.52	21.62	
<i>SD</i>	3.10	3.93	1.35	
Plastic issue importance				$F (1; 77) = 1.80, p = 0.18$
<i>Mean</i>	1.66	1.52	1.82	
<i>SD</i>	1.01	0.70	1.29	
N (consent)	84	45	39	$X^2 (2, N = 88) =$
No consent	4	1	3	2.41, $p = 0.30$

Note. The perception of importance of the issue of plastic was measured on a 7-point scale ranging from (1) “very important” to (7) “not important at all”.

Appendix Chapter 4

4.A. Experimental Scenario

People working in the public sector often have to deal with demands from multiple stakeholders, such as citizens, clients, inspectorates, as well as the central government and its ministries. These demands can sometimes be conflicting, or they can require the performance of very different tasks. Prioritizing the demands of one stakeholder over the others, therefore, becomes necessary.

We would like to ask you to consider the following situation:

Imagine that you work for a public sector organization which provides services to clients. Your organization constantly faces demands from multiple stakeholders.

Two stakeholders have communicated demands to your organization. You have been asked to perform tasks to respond to both demands, however, you can only do one at a time, so you must decide which one to perform first. Which stakeholder's demand would you priorities?

Table 4.A.1 *Example profile choice*

	Stakeholder 1	Stakeholder 2
Stakeholder expertise	<i>Fully understands the type of work your organization performs</i>	<i>Fully understands the type of work your organization performs</i>
Possible consequence for non-compliance with demand	<i>Bad press, damage to your organization's reputation</i>	<i>Financial damage to your organization (fine or budget reduction)</i>
Likelihood of imposing consequence for non-compliance with demand	<i>Almost certainly</i>	<i>Very unlikely</i>
Previous experience with stakeholder	<i>Many struggles with this stakeholder in the past</i>	<i>No struggles at all with this stakeholder in the past, good collaboration</i>

4.B. Sampling procedure for the United Kingdom sample

We sampled using two presets in Prolific.ac: country of residence and employer type. We chose United Kingdom as a country of residence and not nationality since nationality is not a prerequisite for employment in the public sector (in most cases), and because UK nationals can potentially work in the public sector of other countries. We aimed to capture respondents who work in the UK public sector, and thus country of residence is in our opinion more appropriate than nationality for that goal. For employer type we selected local, state and federal government employee. There were 1125 respondents in Prolific's pool that have been active in the last 90 days and matched these characteristics at the time of our sampling. In addition we added two yes/no questions at the start of our survey which asked the participants 1) Do you currently reside in the UK? 2) Are you employed by a governmental or a public sector organisation?. If they answered negatively to at least one of the questions, they were not permitted to proceed with the survey.

4.C. Overview of the surveyed organizations and samples in the Netherlands

Table 4.C.1 *Surveyed organizations*

	Number of employees	Parent ministry	Completed responses	Response rate
Belastingdienst Toeslagen (BD-TL)	~ 28000	Ministry of finance	111	43.5%
Centraal Administratie Kantoer (CAK)	~1250	Ministry of health, welfare and sport	95	63.3%
Dienst Uitvoering Onderwijs (DUO)	~3000	Ministry of education, culture and science	181	38.1%
Uitvoeringsinstituut Werknemersverzekeringen (UWV)	~17000	Ministry of social affairs and employment	216 126 email 90 intranet	79.4% 0.5%

Table 4.C.2 Overview of participant characteristics per organization

	BD-TL		CAK		DUO		UWV	
	N	%	N	%	N	%	N	%
Gender								
Male	42	37.8%	31	32.6%	59	32.6%	93	43.1%
Female	60	54.1%	61	64.2%	106	58.6%	111	51.4%
Education								
Basisonderwijs	0	0.0%	0	0.0%	0	0.0%	1	0.5%
VMBO of MBO 1	11	9.9%	6	6.3%	5	2.8%	6	2.8%
HAVO, VWO of MBO 2-4	55	49.5%	26	27.4%	64	35.4%	28	13.0%
HBO	27	24.3%	34	35.8%	68	37.6%	109	50.5%
Universitaire opleiding	12	10.8%	27	28.4%	32	17.7%	63	29.2%
Manager								
Yes	3	2.7%	12	12.6%	4	2.2%	12	5.5%
No	103	92.8%	81	85.3%	166	92.7%	195	90.3%
Supervising employees								
Yes	18	16.2%	22	23.2%	28	15.5%	45	20.8%
No	87	78.4%	71	74.7%	142	78.5%	161	74.5%
Client interaction								
Very often	45	40.5%	44	46.3%	104	57.5%	92	42.6%
Often	37	33.3%	15	15.8%	40	22.1%	57	26.4%
Sometimes	15	13.5%	6	6.3%	14	7.7%	36	16.7%
Rarely	6	5.4%	8	8.4%	6	3.3%	10	4.6%
Never	3	2.7%	20	21.1%	6	3.3%	12	5.5%
	Mean (SD)	Min / Max						
Age	49.4 (11.0)	20 / 65	42.9 (10.7)	25 / 64	45.3 (10.7)	26 / 66	50.3 (10.6)	25 / 67
Work tenure	15.2 (11.2)	1 / 42	9.4 (5.6)	0.5 / 22	16.0 (10.7)	0 / 37	20.1 (13.0)	0 / 46
Total N	111		95		181		216	

4.D. Validity considerations

Table 4.D.1 Responses to: “Do you recognize the situation provided in this study in your work?”

	UK	NL			All	
		BD-TL	CAK	DUO		UWV
I deal with similar situations very often	64 (10.7%)	8 (7.2%)	25 (26.3%)	13 (7.2%)	14 (6.5%)	60 (10.0%)
I sometimes deal with similar situations	202 (33.7%)	51 (45.9%)	40 (42.1%)	85 (47.0%)	112 (51.9%)	288 (47.8%)
I recognize these situations in the work of my colleagues, but not in my own work	160 (26.7%)	20 (18.0%)	24 (25.3%)	33 (18.2%)	43 (19.9%)	120 (19.9%)
Such situations are very rare in my workplace	117 (19.5%)	9 (8.1%)	2 (2.1%)	24 (13.3%)	26 (12.0%)	61 (10.1%)
I have never heard of, or encountered such a situation before	57 (9.5%)	19 (17.1%)	2 (2.1%)	15 (8.3%)	13 (6.0%)	49 (8.1%)
(No answer)	0 (0.0%)	4 (3.6%)	2 (2.1%)	11 (6.1%)	8 (3.7%)	25 (4.1%)
N	600	111	95	181	216	603

4.E. Overview of sample characteristics post matching

Table 4.E.1 *Overview of the participant characteristics in the matched sample*

	UK sample		Dutch sample	
	N	%	N	%
<i>Gender</i>				
Male	83	26.6%	92	29.5%
Female	229	73.4%	220	70.5%
<i>Education</i>				
Secondary school or lower	17	5.5%	13	4.2%
Post-secondary school qualifications	85	27.2%	85	27.2%
Bachelor's degree or higher	210	67.3%	214	68.6%
<i>Manager</i>				
Yes	34	10.9%	28	9.0%
No	278	89.1%	284	91.0%
<i>Supervising employees</i>				
Yes	82	26.3%	76	24.4%
No	230	73.7%	236	75.6%
<i>Client interaction</i>				
Very often	161	51.6%	151	48.4%
Often	59	18.9%	74	23.7%
Sometimes	42	13.5%	43	13.8%
Rarely	26	8.3%	17	5.4%
Never	24	7.7%	27	8.7%
	Mean (SD)	Min / Max	Mean (SD)	Min / Max
<i>Age</i>	43.0 (10.7)	20 / 71	43.5 (11.0)	20 / 67
<i>Work tenure</i>	11.0 (8.8)	.2 / 41	11.7 (9.4)	0 / 39
Total N	312		312	

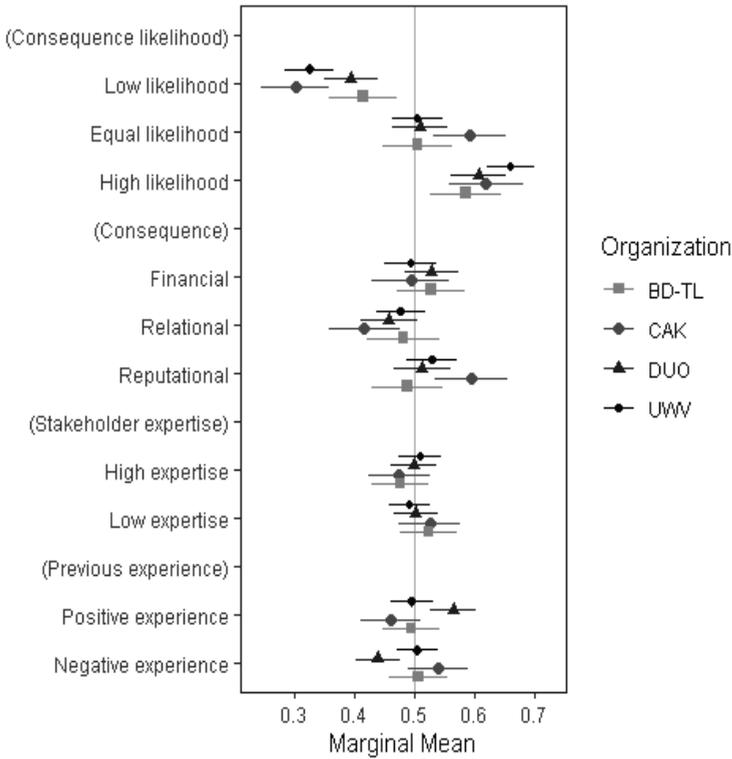
4.F. Organizational effects

Surveyed organizations. In the Netherlands, participants were selected from four executive agencies providing direct financial benefits or grants to citizens. The organizations provide income support (UWV), tax allowances (Belastingdienst-Toeslagen), student loans (DUO), and health care benefits (CAK). These agencies, however, differ importantly in terms of size, as well as the extent to which they are in the public eye. Specifically, CAK is the smallest among the four, and very rarely featured in public discussions. In contrast, BD-TL and UWV are among the largest executive agencies in the Netherlands, and they receive great public attention, although not always positive.

Analysis. Do the general patterns observed for the Dutch civil servants replicate on an organizational level? The analysis of the prioritization decisions for the respondents from the four surveyed Dutch executive agencies are presented in the figure 4.F.1 below. An omnibus F-test points out to some smaller yet significant organizational differences ($F(21; 4556)=3.28, p<0.01$). As it can be observed from the figure, while all organizations respond in a similar way to the threat of sanction – prioritize when the threat is high and de-prioritize when the threat is low – their sensitivity to this threat differs. Thus, we see CAK (*Centraal Administratie Kantoor*) civil servants, for example, a much more risk averse than the civil servants from the other organizations, since they assign higher priority to stakeholders even when the threat of sanction is 50-50 . When it comes to the effects of the different types of sanctions, we observe the general pattern of prioritizing stakeholders who threaten financial and reputational sanction more than ones that threaten a relational sanction. Here again, CAK civil servants stand out somewhat due to the importance they give to the reputational sanction. In terms of stakeholder expertise, we observe a consistent absence of effect. Finally, regarding the effect of previous experience on prioritization

decisions, we observe some variation: DUO employees give priority to stakeholders with whom they have had positive previous experience, CAK has the tendency to do the opposite, while this does not bear any importance for BD-TL and UWV employees.

Figure 4.F.1 *Organizational effects*

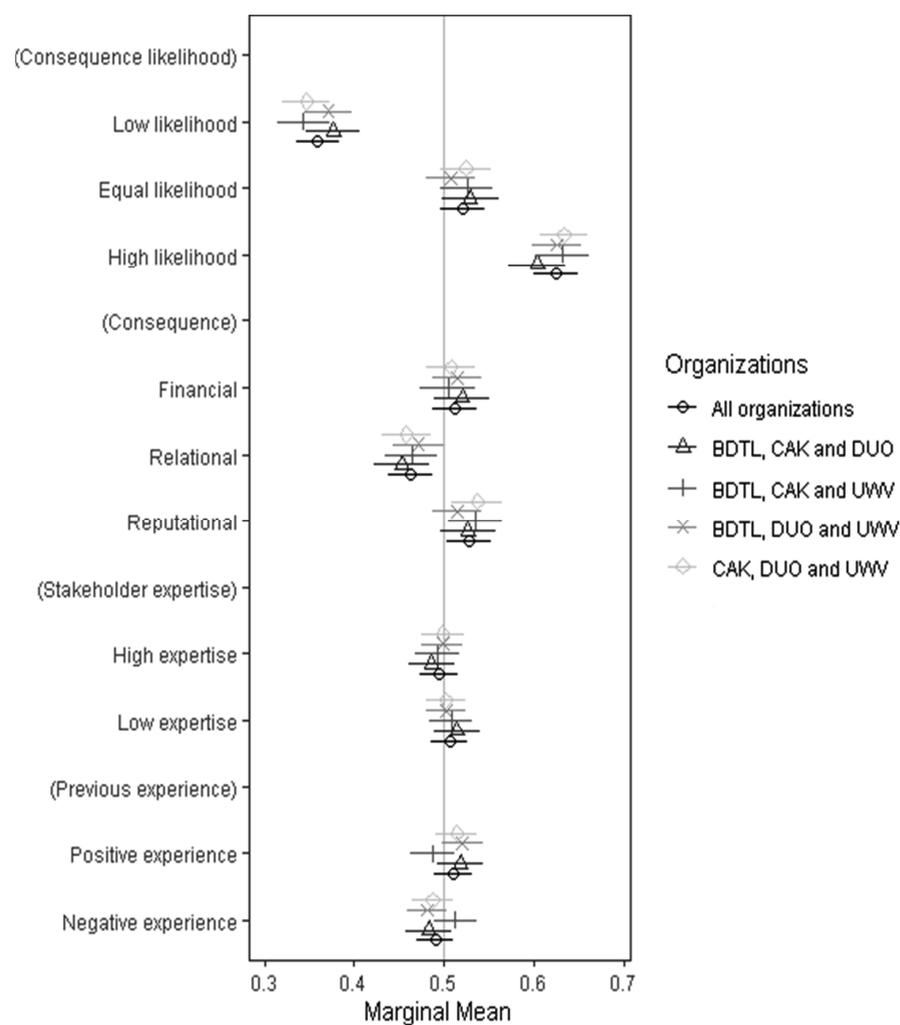


The results again indicate that the effect of sanctions is consistent, although CAK appears to be an outlier. The CAK is an executive agency that carries out various social security regulations for the government and is generally less well-known amongst the general public than the other three agencies. Interestingly, the CAK participants are the most driven by reputational threat in their accountability priorities. These differences in organizational patterns suggest that different

organizational characteristics and experiences might play a role in decisions regarding stakeholder prioritization. Our study, however, is not able to probe further into the sources of these organizational and contextual effects.

In order to assess whether the organizational differences observed here have a strong bearing on the overall results, we conduct a sensitivity analysis whereby we systematically exclude one of the organizations from the overall sample one-by-one, and compare the reduced samples to each other and to the overall sample. In this way, we are able to assess whether a single organizational sample drives the results significantly into a particular direction. The results of this analysis are presented in the figure 4.F.2 below. An omnibus test ($F(28; 18301)=0.80, p=0.76$) indicates that there is no significant difference in the patterns displayed by the five samples.

In sum, while certain organizational specifics and experiences appear to have an effect on how decision-makers within them weigh their considerations in multiple accountability dilemmas, the overall patterns remain largely unaffected. Further research could focus on exploring the relevant organizational experiences and characteristics which bring about differences in how individuals within them respond to multiple accountability dilemmas.

Figure 4.F.2 *Sensitivity analysis*

4.G. Detailed results

Table 4.G.1 Marginal means and standard errors

	All	UK	NL	UK - matched	NL - matched	BD- TL	CAK	DUO	UWV
Consequence likelihood									
<i>Low likelihood</i>	0.38 (0.01)	0.39 (0.01)	0.36 (0.01)	0.42 (0.02)	0.34 (0.02)	0.41 (0.03)	0.30 (0.03)	0.39 (0.02)	0.33 (0.02)
<i>Equal likelihood</i>	0.51 (0.01)	0.51 (0.01)	0.52 (0.01)	0.50 (0.02)	0.51 (0.02)	0.51 (0.03)	0.59 (0.03)	0.51 (0.02)	0.51 (0.02)
<i>High likelihood</i>	0.61 (0.01)	0.59 (0.01)	0.62 (0.01)	0.58 (0.02)	0.65 (0.02)	0.58 (0.03)	0.62 (0.03)	0.61 (0.02)	0.66 (0.02)
Consequence									
<i>Financial</i>	0.53 (0.01)	0.56 (0.01)	0.51 (0.01)	0.54 (0.02)	0.54 (0.02)	0.53 (0.03)	0.49 (0.03)	0.53 (0.02)	0.49 (0.02)
<i>Relational</i>	0.45 (0.01)	0.43 (0.01)	0.46 (0.01)	0.44 (0.02)	0.44 (0.02)	0.48 (0.03)	0.42 (0.03)	0.46 (0.02)	0.48 (0.02)
<i>Reputational</i>	0.52 (0.01)	0.51 (0.01)	0.53 (0.01)	0.52 (0.02)	0.52 (0.02)	0.49 (0.03)	0.59 (0.03)	0.51 (0.02)	0.53 (0.02)
Stakeholder expertise									
<i>High expertise</i>	0.53 (0.01)	0.56 (0.01)	0.49 (0.01)	0.55 (0.01)	0.49 (0.01)	0.48 (0.02)	0.47 (0.03)	0.50 (0.02)	0.51 (0.02)
<i>Low expertise</i>	0.47 (0.01)	0.44 (0.01)	0.51 (0.01)	0.45 (0.01)	0.51 (0.01)	0.52 (0.02)	0.53 (0.03)	0.50 (0.02)	0.49 (0.02)
Previous experience									
<i>Positive experience</i>	0.53 (0.01)	0.55 (0.01)	0.51 (0.01)	0.55 (0.01)	0.52 (0.01)	0.49 (0.01)	0.46 (0.03)	0.56 (0.02)	0.50 (0.02)
<i>Negative experience</i>	0.47 (0.01)	0.45 (0.01)	0.49 (0.01)	0.45 (0.01)	0.48 (0.01)	0.51 (0.02)	0.54 (0.03)	0.44 (0.02)	0.50 (0.02)

Note. The table displays the marginal means and provides their standard errors in brackets.

Table 4.G.2 Linear regression models with clustered standard errors for the pooled sample and the two country samples

	All	UK	NL	UK - matched	NL - matched
Intercept	0.57*** (0.01)	0.49*** (0.02)	0.65*** (0.02)	0.49*** (0.03)	0.67*** (0.03)
Consequence likelihood (ref: High likelihood)					
Equal likelihood	-0.09*** (0.01)	-0.08*** (0.02)	-0.10*** (0.02)	-0.07*** (0.02)	-0.13*** (0.02)
Low likelihood	-0.23*** (0.01)	-0.20*** (0.02)	-0.26*** (0.02)	-0.17*** (0.03)	-0.31*** (0.03)
Consequence (ref: Reputational)					
Relational	-0.07*** (0.01)	-0.08*** (0.02)	-0.07*** (0.2)	-0.08*** (0.03)	-0.09*** (0.02)
Financial	0.02 (0.01)	0.05*** (0.02)	-0.02 (0.02)	0.03 (0.03)	-0.01 (0.03)
Stakeholder expertise (ref: Low expertise)					
High expertise	0.05*** (0.01)	0.12*** (0.01)	-0.01 (0.01)	0.10*** (0.02)	-0.02 (0.02)
Previous experience (ref: Negative experience)					
Positive experience	0.06*** (0.01)	0.09*** (0.02)	0.02 (0.02)	0.10*** (0.02)	0.03 (0.02)
R ²	0.05	0.06	0.05	0.05	0.07
R ² adjusted	0.05	0.06	0.05	0.04	0.07
N observations	4637	2345	2292	1225	1216
N individuals	1203	600	603	312	312

Note. Robust standard errors in parentheses; ***p < .01, **p < .05, *p < .1.

Table 4.G.3 Linear regression models with clustered standard errors for the organizational samples

	BD-TL	CAK	DUO	UWV
Intercept	0.60*** (0.05)	0.78*** (0.05)	0.56*** (0.04)	0.69*** (0.03)
Consequence likelihood (ref: High likelihood)				
Equal likelihood	-0.07* (0.04)	-0.03 (0.04)	-0.10*** (0.03)	-0.15*** (0.03)
Low likelihood	-0.17*** (0.05)	-0.32*** (0.05)	-0.21*** (0.04)	-0.33*** (0.03)
Consequence (ref: Reputational)				
Relational	-0.01 (0.05)	-0.18*** (0.04)	-0.07** (0.03)	-0.05 (0.03)
Financial	0.05 (0.05)	-0.09** (0.04)	0.01 (0.03)	-0.04 (0.03)
Stakeholder expertise (ref: Low expertise)				
High expertise	-0.04 (0.04)	-0.05 (0.04)	0.01 (0.03)	0.01 (0.02)
Previous experience (ref: Negative experience)				
Positive experience	-0.01 (0.04)	-0.08* (0.04)	0.13*** (0.03)	-0.02 (0.02)
R ²	0.02	0.12	0.05	0.07
R ² adjusted	0.02	0.11	0.05	0.07
N observations	423	370	683	816
N individuals	111	95	181	216

Note. Robust standard errors in parentheses; ***p < .01, **p < .05, *p < .1.

4.H. Effects of personal and professional characteristics

Our data includes a range of personal and professional characteristics of the respondents, which could have a potential bearing on the way they make their prioritization choices. Here we report omnibus ANOVA tests on the effect of each characteristic on prioritization choices. A statistically significant omnibus ANOVA test signifies that there is a potentially meaningful difference in the prioritization decisions of the respondents depending on their personal or professional characteristics.

Table 4.H.1 *Effect of personal and professional characteristics*

	All	UK	NL	BD-TL	CAK	DUO	UWV
Personal characteristics							
Gender	F (7; 9042) = 2.81, p = 0.01	F(7; 4660) = 0.38, p = 0.92	F(7; 4368) = 1.40, p = 0.20	F(7; 776) = 0.56, p = 0.79	F(7; 714) = 1.57, p = 0.14	F(7; 1272) = 1.57, p = 0.14	F (7; 1564) = 0.41, p = 0.89
Education	F (14; 9137) = 2.84, p < 0.01	F(14; 4669) = 2.42, p < 0.01	F (14; 4447) = 0.94, p = 0.52	F (14; 793) = 0.67, p = 0.80	F (14; 713) = 1.53, p = 0.09	F (14; 1297) = 2.05, p = 0.01	F (14 ; 1581) = 1.38, p = 0.15
Age	F (14; 8783) = 1.68, p = 0.05	F (14; 4613) = 1.62, p = 0.07	F (14; 4149) = 1.27, p = 0.21	F (14; 755) = 1.32, p = 0.19	F (14; 681) = 0.74, p = 0.74	F (14; 1181) = 1.96, p= 0.02	F (14; 1469) = 0.89, p = 0.57
Professional characteristics							
Manager	F (7; 9152) = 2.55, p = 0.01	F(7; 4668) = 4.96, p < 0.01	F (7; 4470) = 0.92, p = 0.49	F (7; 808) = 2.19, p = 0.03	F (7; 720) = 0.63, p = 0.73	F (7; 1312) = 0.34, p = 0.94	F (7; 1588) = 0.75 , p = 0.63
Supervising employees	F (7; 9144) = 1.07, p = 0.38	F (7; 4676) = 2.02, p = 0.03	F (7; 4454)= 1.46, p = 0.18	F (7; 800) = 1.29, p = 0.25	F (7; 720) = 0.17, p = 0.99	F (7; 1312) = 1.41, p = 0.19	F (7; 1580) = 1.04, p = 0.40
Tenure	F (14; 8799) =	F (14; 4621) =	F (14; 4157) =	F (14; 745) =	F (14; 657) =	F (14; 1179) =	F (14; 1513) =

	2.03, p = 0.01	2.26, p < 0.01	1.33, p = 0.18	0.94, p = 0.52	1.29, p = 0.21	1.82, p = 0.03	0.99, p = 0.45
Client interaction	F (14; 9153) = 1.12, p = 0.34	F (14; 4669) = 0.91, p = 0.54	F (14; 4463) = 0.61, p = 0.86	F (14; 801) = 0.50, p = 0.93	F (14; 713) = 1.13, p = 0.33	F (14; 1305) = 0.75, p = 0.72	F (14; 1581) = 1.30, p = 0.20

Note. Tenure was categorized in three groups: less than 5 years – new employee; 5-20 years – experienced employee; more than 20 years – senior employee. Client interaction was recoded into three categories to facilitate analysis: very often or often, sometimes, rarely or never. To facilitate analysis, we created three age groups: younger than 35, from 35 to 50, older than 50. These three categories have all roughly a range of 15 years and split the respondents in relatively even three groups.

Appendix Chapter 5

5.A. Overview of measures

Table 5.A.1 *Overview of measures used*

Measure	Description	Key sources
<i>High-effort strategies</i>		
Decision-making time	Time to make decision, measured in seconds.	Kassin et al., 1991; Siegel-Jacobs & Yates, 1996
Decision-making complexity	Integrative complexity in thinking displayed in decision justification.	Green et al., 2000; Tetlock, 1983a
Information search	Selecting an option to view more information before making a decision.	Huber et al., 2009; Ten Velden et al., 2010
<i>Low-effort strategies</i>		
Asking for help	How likely to ask a colleague / boss for help. (7-point scale)	Hwang & Han, 2017
Procrastination	How likely to help a colleague with an urgent case and postpone taking an action. (7-point scale)	Green et al., 2000; Tetlock & Boettger, 1994
Buck-passing	How likely to transfer the responsibility for the case to a colleague if possible. (7-point scale)	Green et al., 2000; Tetlock & Boettger, 1994

5.B. Low-effort coping strategies items

Table 5.B.1 *Questions capturing low effort strategies*

Item	Question
<i>Looking for help</i>	How likely is it that you would seek help or advice from a colleague in this situation?
	How likely is it that you would seek help or advice from your supervisor in this situation?
<i>Procrastination</i>	Imagine a colleague comes to you with an urgent request for help. If you decide to help, it will certainly take two full working days. Do you drop everything to help or do you handle the situation of the care home first?
<i>Buck-passing</i>	If you were given the opportunity to immediately transfer the case of the nursing home to a colleague, would you do that?

5.C. Survey distribution details

Table 5.C.1 *Distribution details*

Distribution channel	Sample	Responses
Dutch Health and Youth Care Inspectorate (IGJ) intranet	~ 750	13
Knowledge Platform of the Regulation Academy (Academie voor toezicht kennisplatform)	52	2
Regulation Academy (Academie voor toezicht) email list	244	169
Netherlands School of Public Administration (NSOB) - Learning Workshop Supervision and Compliance alumni	192	58
Newsletter for regulators	8257	18
Linkedin group for regulators	8260	10
Total	Not meaningful - overlap between distributions	270

5.D. Overview of participant characteristics

Table 5.D.1 Participant characteristics per experimental group

	Full sample	Experimental group				Difference between experimental groups
		Single forum compatible demands	Single forum conflicting demands	Multiple forums compatible demands	Multiple forums conflicting demands	
	N	N	N	N	N	
Manager						
yes	38	6	6	13	13	$X^2(3, N = 264) = 2.68, p = 0.44$
no	226	56	46	67	57	
Supervising employees						
yes	110	24	19	36	31	$X^2(3, N = 265) = 1.52, p = 0.68$
no	155	38	34	44	39	
Contact with clients / supervisees						
Very often	104	27	25	24	28	Two-sided Fisher's test $p = 0.28$
Often	79	20	15	28	16	
Sometimes	58	10	11	17	20	
Almost never or never*	24	5	2	11	6	
Supervision experience						
Yes, also specifically with care homes	14	6	2	2	4	Two-sided Fisher's test $p = 0.80$
Yes, also specifically with intramural facilities	24	6	5	6	7	
Yes	220	48	45	68	59	
No experience	12	3	2	5	2	

<i>with supervision</i>						
Highest educational attainment						
<i>VMBO or MBO1 (secondary vocational education)</i>	2	0	1	1	0	Two-sided Fisher's test p = 0.15
<i>HAVO, VWO or MBO 2-4 (secondary education)</i>	14	8	2	3	1	
<i>HBO (higher professional education)</i>	97	20	22	26	29	
<i>University education</i>	152	34	28	50	40	
<i>Gender</i>						
<i>Male</i>	151	41	25	48	37	Two-sided Fisher's test p = 0.30
<i>Female</i>	109	21	26	31	31	
<i>Other / rather not say</i>	5	0	2	1	2	
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
Age	52.33 (9.34)	50.82 (9.74)	52.45 (9.27)	53.28 (9.22)	52.54 (9.20)	F(3, 253) = 0.80, p = 0.49
Professional tenure	11.67 (9.93)	10.67 (10.36)	11.52 (9.86)	12.10 (10.12)	12.18 (9.51)	F(3, 262) = 0.32, p = 0.81
N	270	63	54	81	72	

Note. The categories of almost never and never were collapsed into one due to the low number of respondents in each cell.

5.E. Validity considerations

Table 5.E.1 Responses to: “In this case you were confronted with several demands. Do you recognize such situations in your work?”

	N	%
Yes, I often have to deal with this myself	146	54.07%
Yes, I sometimes have to deal with this	97	35.92%
I recognize this in the work of my colleagues, but I don't have to deal with it myself	23	8.52%
These kinds of situations hardly ever occur in our organization	4	1.48%
I have never dealt with this and never heard about it in our organization	0	0.00%
Test of difference between manipulation groups: Two-sided Fisher's test $p = 0.91$		

5.F. Experimental scenario in Dutch

<p>Scenario</p>		<p><i>Stelt u zich de volgende situatie voor: U bent inspecteur bij de Inspectie Gezondheidszorg en Jeugd. U verricht een inspectie bij een verzorgingshuis na signalen van bezorgde belanghebbenden dat er problemen zijn. Tijdens die inspectie constateert u inderdaad verschillende belangrijke tekortkomingen en zijn er indicaties van mogelijke verwaarlozing van bewoners. U constateert op dit moment echter geen directe negatieve consequenties voor de bewoners. U weet niet zeker wat te doen. Aan de ene kant zijn er duidelijk risico's voor het welzijn van de bewoners in het verzorgingshuis. Aan de andere kant heeft het opleggen van stevige maatregelen mogelijk ook nadelige gevolgen voor de bewoners, omdat het de continuïteit van het verzorgingshuis kan bedreigen, het leven van de kwetsbare ouderen verstoort en dus stressoplevert voor de bewoners en hun familieleden als ook voor de leiding en medewerkers. Er is veel media-aandacht voor deze casus. Als inspecteur heeft u te maken met verschillende relevante belanghebbenden, zoals bewoners en hun families, zorgverleners, bestuurders, de branchevereniging en soms ook de gemeente en het ministerie, die soms in het openbaar uitspreken wat u als inspecteur van de IGJ moet doen. We geven hier onder weer wat er over de situatie is gezegd door belangrijke belanghebbenden. Daarna vragen we u te besluiten wat er volgens u in deze casus moet gebeuren en om dit goed toe te lichten.</i></p>	
		<p>Demand alignment</p>	
		<p>Compatible demands</p>	<p>Conflicting demands</p>
<p>Number of accountability forums</p>	<p>One forum</p>	<p>Belanghebbende A heeft heel nadrukkelijk gesteld dat de inzet van ingrijpende maatregelen tegen het verzorgingshuis geen nadelige gevolgen moet hebben voor de bewoners. Daarom is het belangrijk dat <u>verzorgingshuizen in dit soort gevallen de tijd moeten krijgen om wel aan de kwaliteitseisen te voldoen,</u> gezien de problemen die de</p>	<p>Belanghebbende A heeft in het openbaar heel scherp gesteld dat verwaarlozing van ouderen in geen enkele vorm en onder geen beding kan worden getolereerd. Er moet hard worden opgetreden tegen dit soort verzorgingshuizen waar het welzijn van bewoners op het spel staat. Deze belanghebbende dringt er op</p>

		<p>toepassing van sancties voor bewoners kunnen veroorzaken.</p> <p>Tegelijk heeft dezelfde belanghebbende A gesteld dat de inspectie onmiddellijk een <u>verhoogde frequentie van inspectiebezoeken moet invoeren</u>, om ervoor te zorgen dat het verzorgingshuis zijn problemen oplost.</p>	<p>aan dat dit verzorgingshuis <u>eigenlijk gesloten zou moeten worden of dat een andere serieuze maatregel wordt opgelegd totdat de problemen zijn opgelost</u>.</p> <p>Tegelijkertijd, en in tegenspraak tot het bovenstaande heeft dezelfde belanghebbende A heel nadrukkelijk gesteld dat de inzet van ingrijpende maatregelen tegen het verzorgingshuis geen nadelige gevolgen moet hebben voor de bewoners. Daarom is het belangrijk dat <u>verzorgingshuizen in dit soort gevallen de tijd moeten krijgen om wel aan de kwaliteitseisen te voldoen</u>, gezien de problemen die de toepassing van sancties voor bewoners kunnen veroorzaken.</p>
Two forums		<p>Belanghebbende A heeft heel nadrukkelijk gesteld dat de inzet van ingrijpende maatregelen tegen het verzorgingshuis geen nadelige gevolgen moet hebben voor de bewoners. Daarom is het belangrijk dat <u>verzorgingshuizen in dit soort gevallen de tijd moeten krijgen om wel aan de kwaliteitseisen te voldoen</u>, gezien de problemen die de toepassing van sancties voor</p>	<p>Belanghebbende A heeft in het openbaar heel scherp gesteld dat verwaarlozing van ouderen in geen enkele vorm en onder geen beding kan worden getolereerd. Er moet hard worden opgetreden tegen dit soort verzorgingshuizen waar het welzijn van bewoners op het spel staat. Deze belanghebbende dringt er op aan dat dit verzorgingshuis</p>

		<p>bewoners kunnen veroorzaken.</p> <p>Tegelijk heeft een tweede belanghebbende B gesteld dat de inspectie onmiddellijk een <u>verhoogde frequentie van inspectiebezoeken moet invoeren</u>, om ervoor te zorgen dat het verzorgingshuis zijn problemen oplost.</p>	<p><u>eigenlijk gesloten zou moeten worden of dat een andere serieuze maatregel wordt opgelegd totdat de problemen zijn opgelost.</u></p> <p>Tegelijkertijd, en in tegenspraak tot het bovenstaande heeft een tweede belanghebbende B heel nadrukkelijk gesteld dat de inzet van ingrijpende maatregelen tegen het verzorgingshuis geen nadelige gevolgen moet hebben voor de bewoners. Daarom is het belangrijk dat <u>verzorgingshuizen in dit soort gevallen de tijd moeten krijgen om wel aan de kwaliteitseisen te voldoen</u>, gezien de problemen die de toepassing van sancties voor bewoners kunnen veroorzaken.</p>
--	--	---	--

5.G. Experimental scenario translated to English

Scenario		<p><i>Consider the following situation:</i> You are an inspector at the Health and Youth Care Inspectorate. You perform an inspection at a care home after signals from concerned parties indicating the presence of problems. During the inspection, you indeed find several important shortcomings and indications of possible neglect of the residents. However, you do not notice any direct negative consequences for the residence at the moment. You are not sure what to do. On the one hand, there are clear risks to the well-being of the residents in the care home. On the other hand, the impositions of strict measures may also have adverse consequence for the residents, as it can threaten the continuity of the care home, disrupt the live of the vulnerable elderly, and thus cause stress for the residents and their families as well as for the management and employees. There is a lot of media attention for this case. As an inspector, you have to deal with various relevant stakeholders, such as residents and their families, care providers, administrators, industry associations, and sometimes also the municipality and the ministry, who sometimes publicly state what you should do as an inspector of the Health and Youth Care Inspectorate. Below we provide what has been said about the situation by key stakeholders. We then ask you to decide what you think should be done in this case and to explain this well.</p>	
		Demand alignment	
		Compatible demands	Conflicting demands
Number of accountability forums	One forum	<p>Stakeholder A has stated very emphatically that the use of drastic measures against the care home should not have any negative consequences for the residents. It is therefore important that <u>care homes in these types of situations be given time to meet the quality requirements,</u> given the problems that the imposition of sanctions can cause for the residents.</p>	<p>Stakeholder A has publicly made it very clear that neglect of the elderly cannot be tolerated in any form and under any circumstances. Strong action must be taken against these types of care homes where the well-being of residents is at stake. This stakeholder insists that <u>the care home should be closed or have some other serious measure imposed on it until the problems are resolved.</u></p>

		<p>At the same time, the same stakeholder A has also argued that the inspectorate should <u>immediately introduce an increased frequency of inspection visits</u> to ensure that the care home resolves its problems.</p>	<p>At the same time, and in contradiction to the previous statement, the same stakeholder A has stated very emphatically that the use of drastic measures against the care home should not have any negative consequences for the residents. It is therefore important that <u>care homes in these types of situations be given time to meet the quality requirements</u>, given the problems that the imposition of sanctions can cause for the residents.</p>
	<p>Two forums</p>	<p>Stakeholder A has stated very emphatically that the use of drastic measures against the care home should not have any negative consequences for the residents. It is therefore important that <u>care homes in these types of situations be given time to meet the quality requirements</u>, given the problems that the imposition of sanctions can cause for the residents.</p> <p>At the same time, a second stakeholder B has argued that the inspectorate should <u>immediately introduce an increased frequency of inspection visits</u> to ensure that the care home resolves its problems.</p>	<p>Stakeholder A has publicly made it very clear that neglect of the elderly cannot be tolerated in any form and under any circumstances. Strong action must be taken against these types of care homes where the well-being of residents is at stake. This stakeholder insists that <u>the care home should be closed or have some other serious measure imposed on it until the problems are resolved</u>.</p> <p>At the same time, and in contradiction to the previous statement, a second stakeholder B has stated very emphatically that the use of drastic measures against the care home should not have any negative consequences for the</p>

			residents. It is therefore important that <u>care homes in these types of situations be given time to meet the quality requirements,</u> given the problems that the imposition of sanctions can cause for the residents.
--	--	--	--

Dutch summary

Verantwoordingsmechanismen worden al lange tijd beschouwd als belangrijke instrumenten voor het sturen van gedrag in de publieke sector. Ze worden vaak gezien als een middel om bepaalde normatief wenselijke uitkomsten, zoals het correct uitvoeren van gedelegeerde taken, het voorkomen van machtsmisbruik en het verbeteren van prestaties, te realiseren. Het empirische bewijs over de effecten van verantwoordingsmechanismen in de publieke sector komt echter niet overeen met deze optimistische verwachtingen. Empirisch onderzoek laat zien dat verantwoording vaak geassocieerd met een serie effecten, waarvan vele niet verwacht noch wenselijk zijn.

Dit contrast tussen verwachte en geobserveerde effecten van verantwoording is, in ieder geval deels, het gevolg van een gebrek aan systematisch, empirisch onderzoek naar de causale effecten van verantwoordingsmechanismen in het publieke domein. De verwachte positieve effecten van verantwoording in de bestuurskunde zijn veel sterker onderbouwd door theoretische en normatieve overwegingen dan door empirische bewijslast. Hoewel empirisch bewijs over de beweringen van normatief wenselijke effecten van verantwoording absoluut bestaat, blijven nog zeer veel vragen onbeantwoord. Dit heeft, begrijpelijk, geresulteerd in veel onderzoek naar de mislukkingen, tekortkomingen en problematische resultaten van het toepassen van verantwoordingsmechanismen, die de universele, positieve opvatting over verantwoording ter discussie stellen. Veel van deze onderzoeken focussen echter voornamelijk op specifieke gevallen en voorbeelden van falen, zonder uit te komen in gedeelde kennis of een samenhangende onderzoeksagenda. De onduidelijke conceptuele grenzen van verantwoording en het royale gebruik van dit concept hebben hierbij niet geholpen. Dus terwijl er een grote hoeveelheid theoretisch, normatief, conceptueel en empirisch onderzoek bestaat naar verantwoording, die veel belangrijke inzichten biedt, bestaan er ook nog belangrijke hiaten in het begrip over de consequenties van verantwoording voor besluitvormingsgedrag in de bestuurskunde.

Dit proefschrift beoogt bij te dragen aan het vullen van deze hiaten op te vullen door te focussen op verantwoordingseffecten op het individuele, microniveau in de publieke sector. Dit proefschrift focust specifiek op het besluitvormingsgedrag van ambtenaren met discretionaire bevoegdheden die worden geconfronteerd met verantwoording. Door in de literatuur besproken meso- en macro-effecten van verantwoording door te vertalen naar hun betekenis voor individuele reacties op verantwoording ontstaat inzicht in empirische effecten in het bestuur. Daarnaast maakt een dergelijke op het individu gerichte benadering het mogelijk om gebruik te maken van rijke gedragsmatige en vooral socio-psychologische onderzoekstraditie rond het onderwerp. De overkoepelende onderzoeksvraag die leidend is voor dit onderzoek is: *Hoe beïnvloeden verantwoordingsmechanismen besluitvormingsgedrag in de publieke sector?*

De onderzoeksvraag is vertaald naar vier deelprojecten, welke elk werden geleid door een meer specifieke onderzoeksvraag. De eerste deelvraag, onderzocht in het tweede hoofdstuk, was: *Welke inzichten kunnen uit gedragsonderzoek over verantwoording worden gehaald voor onderzoek in het domein van de bestuurskunde?* Dit deelproject heeft resultaten over causale effecten van verantwoordingsmechanismen in de gedragsliteratuur onderzocht, in relatie tot belangrijke verantwoordingskwesties in de bestuurskundige literatuur. Door middel van een systematische review van 266 experimenten uit de gedragsliteratuur die de effecten van verantwoording op een reeks van uitkomsten onderzochten geïdentificeerd. Het literatuuronderzoek toonde aan dat verantwoording een groot aantal effecten kan hebben op het individuele niveau, waarvan er vele relevant zijn voor kwesties die binnen het openbaar bestuur aan de orde zijn. Het laat in het bijzonder zien dat de druk van verantwoording ervoor zorgt dat individuen een grotere inspanning leveren in besluitvormingssituaties, het ontmoedigt risico-gedrag en vermindert een aantal besluitvormingsbiases. Nog belangrijker, de systematische review laat zien dat

verantwoordingsmechanismen bijzondere nadere kenmerken kunnen hebben die belangrijke consequenties hebben voor hoe ze werken. Dit suggereert dat verantwoording strategisch kan worden gekalibreerd – passend worden gemaakt – voor bepaalde doelen.

De daaropvolgende drie deelprojecten richten zich op de effecten van meer specifieke kenmerken van verantwoordingsmechanismen. Het derde hoofdstuk richt zich op de vraag: *Hoe beïnvloedt de specificiteit van de verantwoordingsnorm besluitvormingsgedrag in de publieke sector?* De resultaten van twee experimentele studies laten echter zien dat het weinig verschil maakte voor deelnemers beslissingen namen of zij verantwoording moesten afleggen over een specifiek geformuleerde, een algemeen geformuleerde of niet-geëxpliciteerde verantwoordingsnorm.

Het derde deelproject, besproken in hoofdstuk 4, gaat dieper in op de invloed van het verantwoordingsforum op het werk van actoren in de publieke sector. Het onderzoek werd gestuurd door de onderzoeksvraag: *Van welke verantwoordingsfora worden eisen geprioriteerd?* Met behulp van een conjoint experiment dat wordt toegepast op twee samples van publieke sector professionals, onderzoekt dit deelproject welke factoren ervoor zorgen dat actoren die zich moeten verantwoorden prioriteit geven aan de eisen van sommige verantwoordingsfora ten koste van die van andere. Het onderzoek laat zien dat de dreiging van sancties stelselmatig de meest belangrijke factor is die prioriteringen bepaalt, terwijl de expertise van het verantwoordingsforum en de relatie met de actor een secundaire, context-afhankelijke rol vervult.

Het vierde en laatste deelproject richt zich op de consequenties die volgen uit het gelijktijdig opereren van meerdere vormen van verantwoording, hetgeen veelvuldig voorkomt in de publieke sector en in de literatuur gewoonlijk wordt aangeduid als het probleem van meervoudige verantwoording. Het deelproject had tot doel om de effecten van het aantal verantwoordingsfora te onderscheiden van de effecten van de onverenigbaarheid van eisen van verantwoordingsfora.

Het werd gestuurd door de onderzoeksvraag: *Hoe beïnvloeden het aantal verantwoordingsfora en de verenigbaarheid van hun eisen het besluitvormingsgedrag in de publieke sector?* Aan de hand van een survey experiment laat dit deelproject zien dat allebei de dimensies van het probleem van meervoudige verantwoording consequenties heeft voor het besluitvormingsgedrag van actoren die zich verantwoorden, in het bijzonder ten aanzien van de waarschijnlijkheid dat zij hulp zoeken en het bespoedigen van hun besluitvorming. Wel is het zo dat de onverenigbaarheid van verantwoordingseisen de sterkte gedragsveranderingen oproepen.

De integratie van theorieën en methoden uit de gedragswetenschap in bestuurskundig onderzoek die dit proefschrift beoogt, leidt tot verschillende overkoepelende bevindingen en conclusies. We concluderen dat de gedragswetenschappelijke inzichten slechts gedeeltelijk succesvol theoretisch kunnen verklaren wat de consequenties zijn van specifieke verantwoordingseffecten in de publieke sector. In aanvulling blijkt dat de methoden die worden gebruikt door gedragswetenschappelijke onderzoekers om de effecten van verantwoording te onderzoeken niet altijd toepasbaar zijn in de context van de publieke sector. Om die reden is de nodige theoretische en methodologische ontwikkeling nodig om nog verder te komen in het begrijpen van de effecten van verantwoordingsmechanismen op het besluitvormingsgedrag van actoren in de publieke sector.

About the author

Marija Aleksovaska was born in 1989 in Skopje. She studied Political Science at the Ss. Cyril and Methodius University in Skopje (2008 – 2012). During this bachelor's program, she studied one year at the University of Nice Sophia Antipolis. After obtaining her bachelor's degree, she pursued a two-year master's program in Political Science at the Central European University in Budapest (2012 – 2014), specializing in political research methodology and social analysis. During her master's studies, Marija was awarded the Departmental Award of Excellence by the Political Science Department at the Central European University, as one of the students with the highest GPA in her cohort. After graduating from Central European University with distinction, she was awarded the Leiden University – CEU Praesidium Libertatis Scholarship which allowed her to pursue a second master's degree in the research master's program in Political Science and Public Administration at Leiden University (2014 – 2016). She completed the second master's program earning a *cum laude*. While pursuing her studies, Marija also worked part time as a researcher in a Skopje-based think-tank, and later worked as a research assistant at Leiden University.

In September 2016 Marija joined the Utrecht University School of Governance as a doctoral researcher working on the NWO-funded project *Calibrating Public Accountability* (CPA). During her PhD trajectory, Marija has published and presented her work in academic journals (Public Administration Review, Public Administration, Journal of Behavioral Public Administration, Public Performance and Management Review, Bestuurskunde) and conferences (EGPA, ECPR, NIG). She has also taken up teaching and supervision responsibilities, and was a member of the PhD council of the Law, Economics and Governance Faculty at Utrecht University. As of February 2021, Marija works as a post-doctoral researcher at Utrecht University School of Governance on the Horizon 2020 project *Trust in Governance and Regulation in Europe* (TiGRE).

A myriad of effects have been attributed to the operation of accountability mechanisms in the public sector. From preventing the abuse of power to improving the effectiveness of public service provision. Many of these suggest causal chains in which accountability mechanisms are expected to stimulate specific behaviors by individuals leading to desirable outcomes in the public domain. Yet, many of these links have been assumed rather than established. By integrating social psychological insights on individual decision-making behavior into the study of public accountability, *Under Watchful Eyes* develops testable expectations of the behavioral consequences of accountability mechanisms in the public sector. Three experimental studies provide reassuring as well as unexpected insights into how accountability pressures shape the decision-making behavior of public sector professionals. These insights will help decision-makers in government and elsewhere to make evidence-based decisions to calibrate and improve existing accountability mechanisms.

