



Accountable for What? The Effect of Accountability Standard Specification on Decision-Making Behavior in the Public Sector

Marija Aleksovka

To cite this article: Marija Aleksovka (2021): Accountable for What? The Effect of Accountability Standard Specification on Decision-Making Behavior in the Public Sector, Public Performance & Management Review, DOI: [10.1080/15309576.2021.1900880](https://doi.org/10.1080/15309576.2021.1900880)

To link to this article: <https://doi.org/10.1080/15309576.2021.1900880>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC



[View supplementary material](#)



Published online: 31 Mar 2021.



[Submit your article to this journal](#)



Article views: 313



[View related articles](#)



[View Crossmark data](#)

Accountable for What? The Effect of Accountability Standard Specification on Decision-Making Behavior in the Public Sector

Marija Aleksovska

Utrecht University

ABSTRACT

This study investigates how civil servants' decision-making behavior is affected by *what* they are held accountable for. We look into the effects of specification of the accountability standard, and analyze the behavioral changes that might arise by holding civil servants accountable for the implementation of specific rules, as opposed to more loosely defined standards, or no predefined standards at all. Drawing on existing debates in public administration, as well as on theoretical accounts from social psychology, we develop two hypotheses outlining advantages and drawbacks of moving to either side of the accountability standard specification continuum. Specifically, we hypothesize a tradeoff between decision-making effort and decision impartiality. We perform our investigation using an online vignette and a classroom experiment. The results from the investigation do not offer clear support for our expectations. The hypothesis suggesting that accountability for general standards has positive effects on decision processes in terms of effort receives some tentative support, while the one linking specific standards to decision impartiality does not. We discuss possible reasons for this outcome, and draw recommendations for further research aiming to integrate psychological insights into public accountability research.


KEYWORDS

behavioral public administration; decision-making; public accountability

Introduction

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,

CONTACT Marija Aleksovska  m.aleksovska@uu.nl  Utrecht University School of Governance, Bijlhouwerstraat 6, 3511 ZC Utrecht, The Netherlands.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15309576.2021.1900880>

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

summarized in four short questions: “Who?”, “To whom?”, “For what?” and “Why?” (Bovens, 2007, 2010; Mulgan, 2003). 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; Romzek & Dubnick, 1987; Romzek et al., 2012), their consequences seem insufficiently investigated and understood (Aleksovska 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.

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). 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 opposed 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 (ABDTOPConsult, 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; Kwon, 2014; Pires, 2011). 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 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.

Theoretical basis for understanding behavior under accountability pressure

Accountability mechanisms in the public domain 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 (Davis et al., 1997; Dicke, 2002; Jos & Tompkins, 2004).

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, 2020).

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 a 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 domain has been emphasized by a number of scholars (for example Hall et al., 2017; Han & Perry, 2020; 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.

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 (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, 2003). 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 been already 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 of 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 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 toward the information considered (Tetlock, 1983; 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 factor 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.

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 (ABD TOPConsult, 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, 2003), 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 considers and weighs different aspects of the problem (Tetlock, 1983; Tetlock et al., 1989). Following our parallel between the knowledge of the forum’s views and the 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.

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., 2009; 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.

Methodological approach

Experimental scenario and design

In order to investigate our outlined hypotheses, we designed a vignette experiment. 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 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 which 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 political debate and polarization. We expected that the salience of the 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 toward 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 (Aleksovska 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 complete experimental scenario and an overview of the steps of the experiment are provided in [appendix¹](#).

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.

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).

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 Suedfeld, 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, 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. 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 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.

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.

Study 1

Data

The experiment was run online, on a panel of public administration student 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 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 toward 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. The results from the manipulation checks are presented in [Table 1](#). We 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. Participants in the control group reported experiencing the highest influence in evaluating the projects, as well as lowest perception of presence of clear rules, 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

Table 1. Manipulation Checks Results from Study 1.

| | 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.

display some differences in these two questions, and in line with our expectations, however, these differences are not statistically significant.

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 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 smaller 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. The results of the analysis on the three measures of decision-making effort are presented in Table 2.

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

Table 2. 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.

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 displayed in Table 2 indicate a pattern according to our expectations, since the group which was not provided rules took the longest to reach a decision, followed by the group with general rules, and the group with specific rules. However, the differences between the groups did not reach statistical significance².

In terms of justification length we observe a slightly different pattern. While the respondents which were not asked to follow rules in their decision-making provided the lengthiest justifications, which is in accordance to our expectations, the ones which were provided general rules provided somewhat shorter justifications than the group provided with specific rules. The differences between the three groups are, however, not statistically significant.

Finally, as predicted, the participants who did not have a clearly predefined accountability standard displayed the highest levels of integrative complexity. The participants who were provided general rules displayed, however, lower degrees of integrative complexity than the ones provided with specific rules, which is contrary to our expectations. The differences between the three groups, however, did not reach statistical significance.

The results of our analysis do show some tendencies which are in line with H1. However, due to the lack of a clear and statistically significant difference between the groups, we cannot conclude that they provide support for our theoretical expectations.

Table 3. Effects of Accountability Standard Specification on Decision Impartiality from Study 1.

| | No rules (control group) | General rules | Specific rules |
|-----------------------------------|--|---------------|----------------|
| Decision consistency | | | |
| Unalikeability coefficient | 0.57 | 0.65 | 0.61 |
| N | 56 | 47 | 48 |
| Summary statistics | $\chi^2 (2, N = 12,034) = 27.54, p < 0.01$ | | |
| Influence of personal preferences | | | |
| Overlap | 34 | 31 | 25 |
| No overlap | 22 | 16 | 23 |
| % Overlap | 60.71 | 65.96 | 52.08 |
| Summary statistics | $\chi^2 (2, N = 151) = 1.94, p = 0.38$ | | |

Decision impartiality

Following H2, we expect that the greater specificity of the accountability standard will result in greater decision impartiality. The analysis of the two measures of decision impartiality is presented in Table 3.

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 higher levels of consistency than the two treatment groups. These differences in decision consistency between the groups were also found to be statistically significant.

The results of the analysis of the influence of personal preferences in the decision-making processes show mixed tendencies. In line with the expectations outlined in H2, the group of participants which were provided with specific rules to follow in their decision-making process had the smallest overlap of personal preferences and decisions made (52.08%). Contrary to our expectations the group provided with general rules to follow in their decision-making process did display higher overlap of personal preferences and decisions made (65.96%) than the group which was not provided any rules (60.71%). The differences between the groups, however, did not reach statistical significance.

The results from the decision consistency measure appear to refute H2, while the ones observed regarding the effect of personal preferences show inconsistent patterns and lack of statistically significant effect. Taken together, they do not provide support for H2.

Discussion

The results of the experiment did not provide robust support for our theoretical expectations, although they did display some of the expected tendencies. 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 integrative 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 by 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.

Study 2

This study employs the same experimental scenario as the first study. Three changes are introduced, however, in order to strengthen the 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. Eighty eight 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 and are 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 [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. The group which was provided with specific rules to guide their decision-making process did think so to a greater extent ($M=2.83$, $SD=1.58$, $N=35$) than the group of participants which were not provided any rules ($M=3.00$, $SD=1.18$, $N=44$). This difference, however, did not reach statistical significance ($F(1; 77)=0.30$, $p=0.58$).

Results

Decision-making effort

We analyze decision making-effort using three measures, as we did in the previous experiment. 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 a 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 4](#). Contrary to our expectations, the

Table 4. 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.

participants provided with specific rules took more time to finish the decision-making task than the participants which were provided no rules. This difference, however, is not statistically significant.

As in the first experiment, we measured the justification length in terms of the number of characters that the respondents wrote in the justification for their decisions. In accordance to our expectations, here 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.

Regarding integrative complexity, we found that the participants which were provided rules on the basis of which they were asked to make their decision displayed somewhat lower levels of integrative complexity than the participants which were not provided with rules. This is in line with our expectations. The difference between the two groups, however, does not reach statistical significance.

We thus observe only tentative support for H1. The results regarding decision-making effort are again mixed. 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, and in fact observe a tendency contrary to our expectations, when it comes to decision-making time. The tendency displayed in the results of integrative complexity are in line with H1, however, they do not reach statistical significance.

Decision impartiality

The results of the analysis on the effect of accountability standard specification on decision impartiality are presented in Table 5. Similarly, like in the

Table 5. Effects of Accountability Standard Specification on Decision Impartiality from Study 2.

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

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 groups 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, the experimental analysis displayed somewhat surprising results, since they are contrary to our expectations outlined in H2. The group of participants provided with specific rules displayed a greater overlap of personal preferences and the decision made (71.43%) than the group with no provided rules (61.36%). The difference in these two groups, however, does not reach statistical significance. Taken together, the results from the two measures of decision impartiality do not offer support for H2.

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.

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 domain 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, some displaying confusing patterns, and others displaying patterns contrary to our expectations. Our results mostly underline the hypothesis making the case for general accountability standards. In both studies, accountability for general standards was mostly related to more decision effort and more integrative complexity, yet the tendencies did mostly not reach significance. Therefore, it might be prudent not to discard this hypothesis just yet, and to subject it 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 its expectations. However, as the majority of the statistical tests did not reach conventional levels of statistical significance, we discuss some possible explanations for this outcome.

First, the unclear results could be due to the design of our experimental tools. These experiments were developed after an extensive survey of the behavioral literature on accountability, and modeled on the basis of previous experimental studies (specifically Tetlock, 1983 and Tetlock et al., 1989). Their theoretical and methodological basis are, thus, firmly rooted in the behavioral literature. The experimental tools, nevertheless, have 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 manipulations might not have been strong enough, leading to insufficient levels of compliance with the experimental instructions.

Another potential limitation of our experimental designs could lie in the operationalization of our core concepts. Specifically, the distinction between our treatment groups with general and specific rules in study 1 might not have

been strong enough, which is exemplified in the relatively close values that the respondents report in the manipulation checks 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 out to the expected effects, for example for integrative complexity, however, they fell short from reaching conventional levels of statistical significance. Therefore, the effects might be there, however, smaller than we anticipated, and therefore not captured by our experiments.

A future study could account for some of the limitations in our experimental designs. 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 ones' 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 experiments 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 these experiments. It is therefore possible that the accountability pressures in the public sector operate through different

mechanisms than the ones captured by our experiments and theoretical model.

Behavioral approaches present a potentially fruitful avenue for understanding the effects of accountability mechanisms in the public sector (Aleksovska 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 domain. 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. 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 modeling of accountability pressures, by evoking accountability relationships as opposed to directly manipulating them. Indirect modeling 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.

Conclusion

This study aimed to provide a step toward 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 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.

Notes

1. Data and online appendix available at: <https://doi.org/10.17026/dans-xka-nure>
2. 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 ($F(2; 148) = 0.17$, $p = 0.84$).

Notes on contributor

Marija Aleksovska is a PhD candidate at the School of Governance, Utrecht University. She investigates the effects of accountability mechanisms on decision-making behavior in the public sector using experimental methods.

References

- ABDTOPConsult. (2019). *Regels en Ruimte Verkenning Maatwerk in dienstverlening en discretionaire ruimte*. ABDTOPConsult.
- 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>
- Arkes, H. R., González-Vallejo, C., Bonham, A. J., Kung, Y. H., & Bailey, N. (2009). Assessing the merits and faults of holistic and disaggregated judgments. *Journal of Behavioral Decision Making*, 23(3), 250–270. <https://doi.org/10.1002/bdm.655>
- Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*, 52(2), 292–306. [https://doi.org/10.1016/0749-5978\(92\)90040-E](https://doi.org/10.1016/0749-5978(92)90040-E)
- Behn, R. D. (2001). *Rethinking democratic accountability*. Brookings Inst 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. <https://doi.org/10.1093/jopart/muy040>

- Boulianne, S., & Basson, D. (2008). Topic saliency. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 892–892). SAGE.
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics*, 33(5), 946–967. <https://doi.org/10.1080/01402382.2010.486119>
- Brennan, G. (1999). Institutionalising accountability: A commentary. *Australian Journal of Public Administration*, 58(1), 94–97. <https://doi.org/10.1111/1467-8500.00078>
- Bringselius, L. (2017). *Tillitsbaserad styrning och ledning: Ett ramverk*. Tillitsdelegationen.
- Brodikin, E. Z. (1997). Inside the welfare contract: Discretion and accountability in state welfare administration. *Social Service Review*, 71(1), 1–33. <https://doi.org/10.1086/604228>
- Busuioc, M., & Lodge, M. (2017). Reputation and accountability relationships: Managing accountability expectations through reputation. *Public Administration Review*, 77(1), 91–100. <https://doi.org/10.1111/puar.12612>
- Carpenter, D. P., & Krause, G. A. (2012). Reputation and public administration. *Public Administration Review*, 72(1), 26–32. <https://doi.org/10.1111/j.1540-6210.2011.02506.x>
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., De Rosario, H., De Rosario, M. H. (2020). Package ‘pwr’. R package version, 1(3).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.
- Conway, L. G., Conway, K. R., Gornick, L. J., & Houck, S. C. (2014). Automated integrative complexity. *Political Psychology*, 35(5), 603–624. <https://doi.org/10.1111/pops.12021>
- 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. <https://doi.org/10.5964/jspp.v8i2.1307>
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.
- Daly, A. J. (2009). Rigid response in an age of accountability: The potential of leadership and trust. *Educational Administration Quarterly*, 45(2), 168–216. <https://doi.org/10.1177/0013161X08330499>
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). Toward a stewardship theory of management. *Academy of Management Review*, 22(1), 20–47. <https://doi.org/10.5465/amr.1997.9707180258>
- 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. <https://doi.org/10.1037/0022-3514.89.3.345>
- 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. <https://doi.org/10.1177/027507402237870>
- Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, 28(3), 376–417.
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly*, 59(3), 237–254. <https://doi.org/10.2307/2787021>
- Gailmard, S. (2014). Accountability and Principal–Agent Theory. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 90–105). Oxford University Press.

- 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. [https://doi.org/10.1016/0030-5073\(83\)90122-8](https://doi.org/10.1016/0030-5073(83)90122-8)
- 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. <https://doi.org/10.1002/job.2052>
- Han, Y., & Perry, J. L. (2020). Conceptual Bases of Employee Accountability: A Psychological Approach. *Perspectives on Public Management and Governance*, 3(4), 340–340. <https://doi.org/10.1093/ppmgov/gvaa014>
- 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. <https://doi.org/10.1111/1467-9566.00291>
- Hoecht, A. (2006). Quality assurance in UK higher education: Issues of trust, control, professional autonomy and accountability. *Higher Education*, 51(4), 541–563. <https://doi.org/10.1007/s10734-004-2533-2>
- Houck, S. C., Conway, I. I. I., Gornick, L. J. (2014). Automated integrative complexity: Current challenges and future directions. *Political Psychology*, 35(5), 647–659. <https://doi.org/10.1111/pops.12209>
- Hupe, P., & Hill, M. (2007). Street-level bureaucracy and public accountability. *Public Administration*, 85(2), 279–299. <https://doi.org/10.1111/j.1467-9299.2007.00650.x>
- 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. <https://doi.org/10.1177/0095399704263479>
- Kader, G. D., & Perry, M. (2007). Variability for categorical variables. *Journal of Statistics Education*, 15(2), 1–16. <https://doi.org/10.1080/10691898.2007.11889465>
- Klimoski, R. J. (1972). The effects of intragroup forces on intergroup conflict resolution. *Organizational Behavior and Human Performance*, 8(3), 363–383. [https://doi.org/10.1016/0030-5073\(72\)90056-6](https://doi.org/10.1016/0030-5073(72)90056-6)
- Koonce, L., Anderson, U., & Marchant, G. (1995). Justification of decisions in auditing. *Journal of Accounting Research*, 33(2), 369–384. <https://doi.org/10.2307/2491493>
- Kwon, I. (2014). Motivation, discretion, and corruption. *Journal of Public Administration Research and Theory*, 24(3), 765–794. <https://doi.org/10.1093/jopart/mus062>
- 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. [https://doi.org/10.1016/S0148-2963\(98\)00067-8](https://doi.org/10.1016/S0148-2963(98)00067-8)
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- 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. <https://doi.org/10.1093/oxfordjournals.jpart.a024272>

- Ministerie van Binnenlandse Zaken En Koninkrijksrelaties. (2020). *Kabinetsreactie Regels en ruimte – Verkenning Maatwerk in dienstverlening en discretionaire ruimte*. Rijksoverheid.
- Molander, A., Grimen, H., & Eriksen, E. O. (2012). Professional discretion and accountability in the welfare state. *Journal of Applied Philosophy*, 29(3), 214–230. <https://doi.org/10.1111/j.1468-5930.2012.00564.x>
- 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. (2003). *Holding power to account: Accountability in modern democracies*. Palgrave.
- O'Donnell, G. A. (1998). Horizontal accountability in new democracies. *Journal of Democracy*, 9(3), 112–126. <https://doi.org/10.1353/jod.1998.0051>
- Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2), 200–249. <https://doi.org/10.1086/517935>
- Ordóñez, L. D., Benson, L., III, & 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. <https://doi.org/10.1006/obhd.1999.2823>
- Overman, S. (2020). Aligning accountability arrangements for ambiguous goals: the case of museums. *Public Management Review*, Ahead of print, 1–21. <https://doi.org/10.1080/14719037.2020.1722210>
- Perry, M., & Kader, G. (2005). Variation as unalikeability. *Teaching Statistics*, 27(2), 58–60. <https://doi.org/10.1111/j.1467-9639.2005.00210.x>
- 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. <https://doi.org/10.1111/j.1748-5991.2010.01083.x>
- Pollitt, C. (2003). *The essential public manager*. McGraw-Hill Education (UK).
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford 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. <https://doi.org/10.1177/0020852300661004>
- Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the public sector: Lessons from the Challenger tragedy. *Public Administration Review*, 47(3), 227–238. <https://doi.org/10.2307/975901>
- 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. <https://doi.org/10.1111/j.1540-6210.2011.02547.x>
- 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. <https://doi.org/10.1080/01900692.2012.661165>
- Schillemans, T. (2016). Calibrating public sector accountability: Translating experimental findings to public sector accountability. *Public Management Review*, 18(9), 1400–1420. <https://doi.org/10.1080/14719037.2015.1112423>
- Schillemans, T., & Bjurström, K. H. (2020). Trust and verification: balancing agency and stewardship theory in the governance of agencies. *International Public Management Journal*, 23(5), 650–676. <https://doi.org/10.1080/10967494.2018.1553807>
- 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. <https://doi.org/10.1093/jopart/muu024>
- 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. <https://doi.org/10.1016/j.jesp.2006.05.010>
- Shankar, P. G., & Tan, H. T. (2006). Determinants of audit preparers' workpaper justifications. *The Accounting Review*, 81(2), 473–495. <https://doi.org/10.2308/accr.2006.81.2.473>
- 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. <https://doi.org/10.1006/obhd.1996.0001>
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, 51(3), 416–446. [https://doi.org/10.1016/0749-5978\(92\)90020-8](https://doi.org/10.1016/0749-5978(92)90020-8)
- Suedfeld, P. (2010). The cognitive processing of politics and politicians: Archival studies of conceptual and integrative complexity. *Journal of Personality*, 78(6), 1669–1702. <https://doi.org/10.1111/j.1467-6494.2010.00666.x>
- 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. <https://doi.org/10.1080/15309576.2016.1177560>
- Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, 45(1), 74–83. <https://doi.org/10.1037/0022-3514.45.1.74>
- Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 331–376). Academic Press.
- Tetlock, P. E. (1999). Accountability theory: Mixing properties of human agents with properties of social systems. In L. L. Thompson, J. M. Levine, & D. M. Messick (Eds.), *Shared cognition in organizations: The management of knowledge* (pp. 117–137). Lawrence Erlbaum Associates Publishers.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109(3), 451–471. <https://doi.org/10.1037/0033-295x.109.3.451>
- 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. <https://doi.org/10.1037//0022-3514.57.4.632>
- 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. <https://doi.org/10.1111/1475-679X.00034>
- 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. <https://doi.org/10.1093/oxfordjournals.jpart.a024377>
- Weldon, E., & Gargano, G. M. (1988). Cognitive loafing: The effects of accountability and shared responsibility on cognitive effort. *Personality & Social Psychology Bulletin*, 14(1), 159–171. <https://doi.org/10.1177/0146167288141016>
- Yang, K. (2012). Further understanding accountability in public organizations: Actionable knowledge and the structure–agency duality. *Administration & Society*, 44(3), 255–284. <https://doi.org/10.1177/0095399711417699>