Research Synthesis

Lessons from five decades of experimental and behavioral research on accountability:
A systematic literature review

Marija Aleksovska*, Thomas Schillemans*, Stephan Grimmelikhuijsen*

Abstract: The study of accountability in public administration has developed largely in parallel to the study of accountability in the behavioral sciences. In an effort to bridge this divide, we present a systematic review of the experimental literature on accountability in the behavioral sciences and draw lessons for public administration. We summarize the findings of 266 experiments exploring the effects of accountability mechanisms, presented in 211 articles published between 1970 and 2016. These findings are organized in four broad themes: effects of accountability on decision-making, behavior, and outcomes; and effects of the specific characteristics of accountability mechanisms. The review shows numerous desirable effects of accountability on individual decision-making and behaviors. This is of high relevance to public administration studies on accountability as it sheds light on causal mechanisms and allows for a balanced perspective on positive and negative effects of various types of accountability mechanisms. It is however not always possible to translate findings from behavioral research directly to public administration settings. We discuss the meaning and value of our findings for public administration studies and develop an agenda for future behavioral research on public sector accountability.

Keywords: Accountability, Experiments, Systematic review, Decision-making

Supplements: [Open materials](#)

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 domain (Bovens, Goodin, & Schillemans, 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; Halachmi, 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, 2012; Yang, 2012). Moreover, several scholars have pointed out that the causal chains through which accountability

* Utrecht University School of Governance
Address correspondence to Marija Aleksovska at (m.aleksovska@uu.nl)

mechanisms operate have not been sufficiently investigated (Brandsma, 2014; Koch & Wüstemann, 2014; Schillemans, 2013; Yang, 2012; Yang, 2014). Thus, we do not yet have a 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 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 with the study of accountability in public administration and 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, Frink, & Buckley, 2017; Harari & Rudolph, 2017; Lerner & Tetlock, 1999). While there is a long tradition of using experimentation in the behavioral sciences, public administration scholars have less experience with this method (Bouwman & Grimmelikhuijsen, 2016; James, Jilke, & Ryzin, 2017), even with the recent rise in interest in behavioral public administration (Grimmelikhuijsen, Jilke, Olsen, & Tummers, 2017). Thus, in this review we provide a comprehensive overview of 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 & 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 & 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 & 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 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 public administration literature and draws conclusions for the field of public administration.

We argue that findings from the behavioral sciences should be integrated into 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, Brewer, Lee, Petrovsky, & van Witteloostuijn, 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.

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, 2014; 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.

Increased 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, Van Thiel, Bouckaert, & Lægreid, 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 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, Schillemans, & 't Hart, 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". They 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, Küebler, & Marcinkowksi, 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 for 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.

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, Denyer, & Smart, 2003). In particular, the search terms "accountability" and "experiment" were used to identify 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.[1] 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.[2] 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.[3] These articles where then closely examined in order to determine whether they meet the two selection criteria described below. 211 articles, presenting 266 unique experiments satisfied the two criteria and formed the basis of 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 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.[4] 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 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 in countries which are not western developed democracies. But while these studies come from a variety of disciplines, they seem
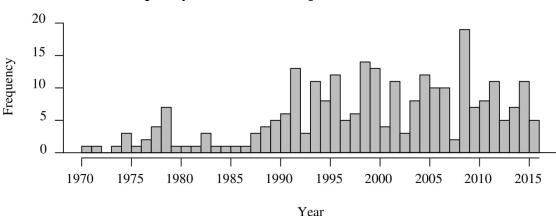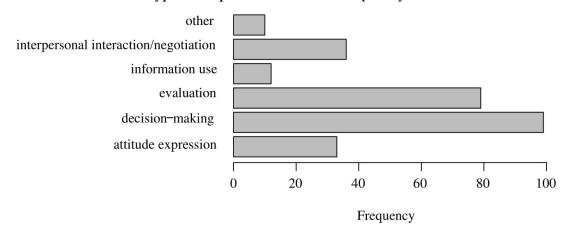
Figure 1
Frequency of Conducted Experiments Over Time

Figure 2
Type of Experimental Task Frequency

to be performed in a fairly similar setup: in the classrooms of campuses of western universities.

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 shows, decision-making and evaluation tasks dominate 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.

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

6

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 a 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. The characteristics of the accountability mechanisms, however, differ across groups. By comparing the results of the experimental groups, these studies show how the characteristics 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 into 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.
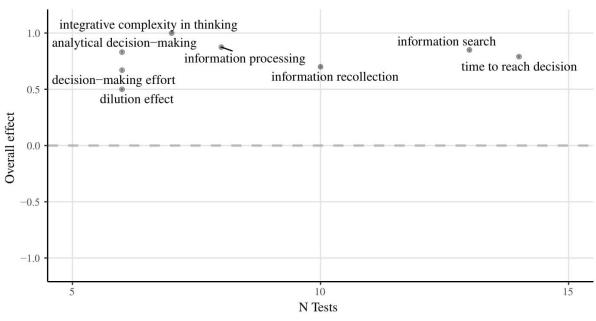
Figure 3
Effects of Accountability on Decision-Making

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).

Accountability and decision-making

As decision-making is at the heart of public administration (Barnard & 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 3 summarizes the central findings of this literature, by presenting the effects that have been tested in at least five different experiments.[5] 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 4 and 5 as well).

The experimental studies investigating the effect of accountability on decision-making consistently found that decision-makers facing accountability, searched for (Huneke, Cole, & Levin, 2004; Lee, Herr, Kardes, & Kim, 1999) 6 and remembered (De Dreu, Beersma, Stroebe, & Euwema, 2006; Roch, 2006; Tetlock, 1983b) more information than decision-makers not facing accountability. They also engaged in deeper information processing (Hess, Rosenberg, & Waters, 2001; Thompson, Roman, Moskowitz, Chaiken, & Bargh, 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, Visser, & Tetlock, 2000; Tetlock, Skitka, & Boettger, 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 4.

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, Sedatole, & Towry, 2005) and motivate actions which enhance the wellbeing of groups (Barreto & Ellemers, 2000).

The effects of accountability on competitiveness and leniency[7] in our summary of the results are
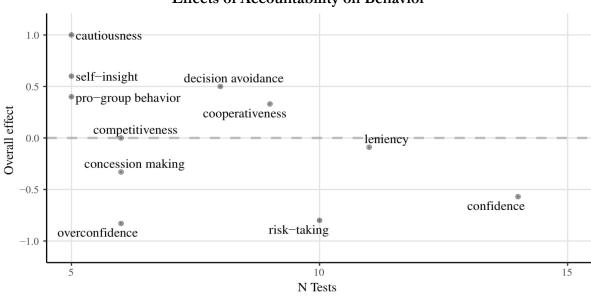
Figure 4
Effects of Accountability on Behavior

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, Schillemans, & 't Hart (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 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 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, Koole, & Steinel, 2000; Scholten, Van Knippenberg, Nijstad, & De Dreu, 2007; Schul & Mayo, 2003), improves judgement consistency (Ashton, 1992; Hagafors & Brehmer, 1983; Ordóñez, Benson, & Beach, 1999), and improves task performance (Frink & Ferris, 1999; Holt, Kinchin, & Clarke, 2012; Tan & Kao, 1999). Accountability also produces positive effects in the social domain, as the results indicate. In particular, accountability pressures have been found to lead to greater benefit to others (Schoenrade, Batson, Brandt, & Loud, 1986), greater contributions to the public good (Lierl, 2016), as well as greater joint outcomes in negotiated agreements (De Dreu et al., 2006).

Thus, these findings consistently show that accountability promotes decisions and performance of a 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.
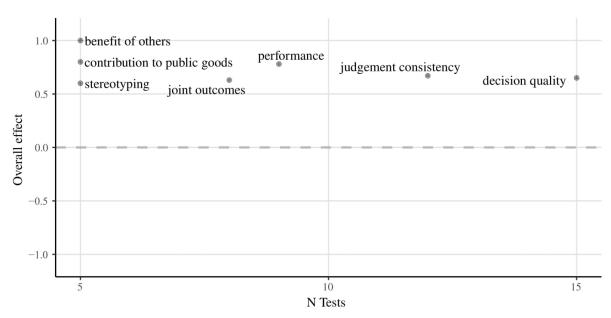
Figure 5
Effects of Accountability on Decision-Making and Behavior Outcomes

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.

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 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.

As Table 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, Mero, & Goodman, 2007).

Table 1
Studied Characteristics of Accountability Mechanisms

Contrasts of accountability mechanisms characteristics	Number of experiments	Number of tested effects
Timing		
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
Standard		
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
Relationship		
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
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

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, Guidice, & Brownlee, 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, Herbst, Har-

din, & Dardis, 2002), and when the forum is considered to be ingroup, as opposed to outgroup[8] (Haccoun & Klimoski, 1975), as well as when the actor is an individual, as opposed to a group (Kroon, 't Hart, & Van Kreveld, 1991).

In sum, these findings show that the characteristics of an accountability mechanism can have a large 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 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.

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.

Simply applying experimental findings to the public sector, however, 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 using experimentation to research 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). This presupposes that effects have been tested in several experiments which limits the scope and focuses attention. We have not chosen to perform a meta-analysis in this study, as we aimed to harvest the experimental evidence on the effects of accountability relating to the main concerns in public administration research, which may include relevant effects found in single studies. For future studies, a thorough meta-analysis of the main experimental effects from an inside-out perspective 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 possible publication bias in such research. 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).

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, for instance in terms of 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 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 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 assess the extent to which 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

Finally, 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, and the relationship between the actor and the forum. This finding suggests 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), and 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). We, however, have little understanding of 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 on the design elements of the accountability mechanisms, 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.

Acknowledgements

Notes

1. Complete search log and a PRISMA flow diagram are available in Appendix 1.
2. These articles are Lerner and Tetlock (1999); Hall, Frink, and Buckley (2017) and Harari and Rudolph (2017).
3. Only one relevant book chapter was found, as well as three dissertations.
4. The complete list of studies included in this review is available in Appendix 2.
5. A detailed representation of all effects is provided in Appendix 3, organized by theme.
6. 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 2.
7. 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.

8. 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.

References

Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*, *52*(2), 292-306.

Barnard, C., & Simon, H. A. (1947). *Administrative behavior: A study of decision-making processes in administrative organization*. New York: Free Press.

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*. Washington DC: Brookings Institution Press.

Bouwman, R., & Grimmelikhuijsen, 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.

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: Oxford University Press.

Bovens, M., Schillemans, T., & Hart, P. T. (2008). Does public accountability work? An assessment tool. *Public Administration*, *86*(1), 225-242.

Brandsma, G. J. (2014). Quantitative analysis. *The oxford handbook of public accountability*. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 143-158). Oxford: Oxford University Press.

Brennan, G. (1999). Institutionalizing accountability: A commentary. *Australian Journal of Public Administration*, *58*(1), 94-97.

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.

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 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.

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.). (2014). *Accountable governance: Problems and promises*. London and New York: Routledge.

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. (2014). The future and relevance of accountability studies. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 661-672). Oxford: Oxford University Press.

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.

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.

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.

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.

Halachmi, A. (2002). Performance measurement, accountability, and improved performance. *Public Performance & Management Review*, *25*(4), 370-374.

Halachmi, A. (2014). Accountability overloads. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 560-573). Oxford: 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.

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.

Hasler, K., Küebler, D., & Marcinkowksi, 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.

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.

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.

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*. London: Sage.

James, O., Jilke, S. R., & Van Ryzin, G. G. (Eds.). (2017). *Experiments in public management research: Challenges and contributions*. Cambridge: Cambridge 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.

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.

Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, *31*(2), 231-245.

Klijn, E. H., & Koppenjan, J. F. (2014). Accountable networks. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 242-257). Oxford: Oxford University Press.

Koch, C., & Wüstemann, J. (2014). Experimental analysis. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 127-142). Oxford: Oxford University Press.

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., 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.

Lægreid, P. (2014). Accountability and new public management. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 324-338). Oxford: Oxford University Press.

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., & 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.

Lierl, M. (2016). Social sanctions and informal accountability: Evidence from a laboratory experiment. *Journal of Theoretical Politics*, *28*(1), 74-104.

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.

Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. Cambridge: Cambridge University Press.

Mulgan, R. (2014). Accountability deficits. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 545-559). Oxford: Oxford University Press.

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.

Palumbo, A. (2017). *Political accountability*. London: 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.

Peters, B. G. (2014). Accountability in public administration. In Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 211-225). Oxford: Oxford University Press.

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., & Grimmelikhuijsen, 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.

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., LeRoux, K., & Blackmar, J. M. (2012). A preliminary theory of informal accountability among network organizational actors. *Public Administration Review*, *72*(3), 442-453.

Schillemans, T. (2008). Accountability in the shadow of hierarchy: The horizontal accountability of agencies. *Public Organization Review*, *8*(2), 175-194.

Schillemans, T. (2011). Does horizontal accountability work? Evaluating potential remedies for the accountability deficit of agencies. *Administration & Society*, *43*(4), 387-416.

Schillemans, T. (2013). *The public accountability review: A meta-analysis of public accountability research in six academic disciplines* (Working paper). Retrieved from Utrecht University Repository https://dspace.library.uu.nl/handle/1874/275784

Schillemans, T. (2016). Calibrating public sector accountability: Translating experimental findings to public sector accountability. *Public Management Review*, *18*(9), 1400-1420.

Schillemans, T., & Busuioc, M. (2014). Predicting public sector accountability: From agency drift to forum drift. *Journal of Public Administration Research and Theory*, *25*(1), 191-215.

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.

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.

Strøm, K. (2000). Delegation and accountability in parliamentary democracies. *European Journal of Political Research*, *37*(3), 261-290.

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.

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., & 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.

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.

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.

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.

Verhoest, K., Van Thiel, S., Bouckaert, G., & Lægreid, P. (Eds.). (2012). *Government agencies: Practices and lessons from 30 countries*. Springer.

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*.

Weigold, M. F., & Schlenker, B. R. (1991). Accountability and risk taking. *Personality and Social Psychology Bulletin*, *17*(1), 25-29.

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 Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.), *The Oxford handbook public accountability* (pp. 159-176). Oxford: Oxford University Press.