Marija Aleksovska
Thomas Schillemans
Stephan Grimmelikhuijsen
Utrecht University

Research Article

Management of Multiple Accountabilities Through Setting Priorities: Evidence from a Cross-National Conjoint Experiment

Marija Aleksovska is a doctoral candidate at the Utrecht School of Governance of Utrecht University. She is interested in the behavioral effects of accountability in the public sector.
Email: m.aleksovska@uu.nl

Thomas Schillemans is a Professor in accountability, behavior, and governance at the Utrecht School of Governance of Utrecht University. His research focuses on the interactions of public sector organizations with various relevant stakeholders from their environment. He specializes in public accountability, public sector governance, trust & control, and the role of the media in policy implementation.
Email: t.schillemans@uu.nl

Stephan Grimmelikhuijsen is an associate professor at the Utrecht School of Governance of Utrecht University. Core themes in his research are: transparency, citizen trust in government, technology and government, experimental methodology, behavioral public administration.
Email: s.g.grimmelikhuijsen@uu.nl

Abstract: *Public sector actors are continuously being held accountable by a multitude of accountability forums. Responding to the forums' demands often requires prioritizing between them. This study investigates how those prioritization choices are made. Drawing on two competing perspectives: the classical view of accountability as "answerability" which emphasizes hierarchy and control, and the modern interpretation of accountability as "management of expectations" which highlights the strategic management of relations, we identify four factors whose influence on prioritization choices we investigate. Using a conjoint experiment, we investigate the prioritization decisions of civil servants in the United Kingdom and the Netherlands. We find that the threat of sanction, which is central in the answerability perspective, is consistently the most important driver of prioritization decisions. The management of expectations, focusing on forum expertise and relationships with the accountability forums, appears to be largely context dependent and helps to explain additional, more fine-grained variations.*

Evidence for Practice
- Managing multiple and conflicting accountability demands is commonplace in the work of civil servants at all levels.
- To successfully resolve multiple accountability dilemmas, civil servants try to reconcile them or give priority to some over others.
- There is a hierarchy of considerations when it comes to solving multiple accountability dilemmas, and the avoidance of sanctions, particularly material and reputational, takes primacy over all other considerations.
- The demands of stakeholders which are seen as highly knowledgeable and with which there is a history of positive working relationship are generally given priority, although the extent to which varies across contexts.
- The threat of sanction is the most powerful tool account holders can employ to ensure that their demands are given priority by civil servants, although softer tools of influence like professional expertise and good professional relationships can have meaningful effects in particular contexts too.

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

Although the challenge of multiple accountabilities is acknowledged widely in the academic literature (Koppell 2005; Schillemans and Bovens 2011), there is very little systematic knowledge about how individual behaviors and decisions in public administration are shaped by the necessity to simultaneously give account to multiple stakeholders (Aleksovska, Schillemans, and Grimmelikhuijsen 2019; Busuioc and Lodge 2017; Yang 2012). Particularly, the question what factors explain why decision-makers at the top or further down the line in the organization prioritize the demands of one stakeholder over others remains unanswered (Busuioc and Lodge 2017, 2; Romzek and Dubnick 1987, 235; Schillemans and

Bovens 2011, 6; Schillemans 2015, 438). Addressing this research gap is not only relevant from an academic perspective, but also from the perspective of governance and regulation, as understanding the behavior of regulatees when confronted with multiple (accountability) demands, is crucial to effective regulation and enforcement (Ashworth, Boyne, and Walker 2002; Boyne, Day, and Walker 2002).

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

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

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

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

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

We discuss the implications for research but also for practices of governance and regulation.

The Unavoidable Multiple Accountabilities: Overview of the Literature

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

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

The influence that multiple accountability pressures have on the work of various types of civil servants has often been characterized as problematic (Schillemans and Bovens 2011). The presence of multiple stakeholders with varying demands creates complexities in the work of the individual decision-maker, since accounting for all of them entails the simultaneous adoption of multiple roles (Ewert 2018; Shortell et al. 1998). Moreover, the various stakeholder demands can come into conflict, creating dilemmas for the decision-maker as to what is the best course of action (Thomann, Hupe, and Sager 2018). The multiple, conflicting accountability demands have been linked to a number of dysfunctions in the public sector, including cynical blame games, transaction and opportunity costs, organizational paralysis, and diminished efficacy (Koppell 2005; Romzek and Dubnick 1987; Schillemans and Bovens 2011).

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

This identified repertoire of strategies for dealing with multiple accountabilities, however, has not been investigated in much

detail. Thus, we do not know when decision-makers decide to just implement all accountability demands that come their way and develop a multiple accountability disorder (Koppell 2005), nor do we know which principles and factors guide prioritization and compromising of demands. The causal links through which the multiple accountability pressures, and their particularities, drive the behavior of public sector decision-makers which ultimately result in some of the suggested outcomes have not been established. This is partly due to the fact that the behavioral responses to multiple accountability pressures have been studied almost exclusively using case studies, with observational or interview data, and this body of literature has not evolved yet into an integrated and developed research agenda (Aleksovska, Schillemans, and Grimmelikhuijsen 2019; Yang 2012).

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

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

Accountability as Answerability
The answerability perspective represents the classical, hierarchical view on governance and places great emphasis on the control and oversight aspect of accountability. Accountability is seen as a tool for ensuring that the behavior of public sector decision-makers complies with the expectations set by accountability forums, and that the decision-maker is answerable for fulfilling its mandate (Aberbach and Rockman 1997; Acar, Guo, and Yang 2008; O'Loughlin 1990). Accountability pressure is seen as external to the public sector decision-maker and originating from the authority of the accountability forum (Willems and van Dooren 2012, 1014). The relationship between the two actors is unequal, and the accountability forum holds power over the public sector decision-maker. This view thus presumes clear principal-agent relationship between the two actors (Aberbach and Rockman 1997; Gailmard 2014).

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

Likelihood of Imposing Sanction. In theory, accountability forums have the possibility to impose a formal or informal sanction to steer

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

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

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

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

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

Accountability as Management of Expectations
The modern management of expectations perspective considers accountability to be more than just answerability. It emphasizes the agency of the individual to formulate strategic responses to the different accountability expectations it faces. Here, control mechanisms are seen as only one, and not necessarily even the most important, elements among a larger set of factors that shape the behavior of decision-makers (Acar, Guo, and Yang 2008; Romzek and Dubnick 1987). The values and interests of the decision-maker take a prominent place in this perspective and act as filtering

mechanisms to the various accountability demands (Busuioc and Lodge 2017). As a result, in the management of expectations perspective, accountability is primarily seen as an internal, rather than external, force (Willems and van Dooren 2012). It represents self-regulation in the face of external expectations through internal values and strategic considerations.

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

Building a good reputation, as well as collaborative relations with accountability forums is thus of key importance for the successful management of multiple accountabilities in these theoretical perspectives. However, not all forums are equally valuable for the reputational and strategic goals of public sector actors. Gaining the approval of the most highly regarded forums would be more beneficial for building esteem, and investing in relations which already have a collaborative nature or potential to do so would be considered more strategic. We thus argue that the forum's expertise and the quality of the relationship with the accountability forum will have an influence on prioritization choices in multiple accountability dilemmas.

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

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

Relationship with the Forum. Many accountability relationships in the public sector have a long-term character. For that reason, building up and maintaining a rapport with the accountability forum is important as it facilitates the continuous management of

their expectations (Lieberherr and Thomann 2019). As Schillemans (2015, 438) finds in his interviews with Dutch public sector managers, the active investment in the relationships with the accountability forums is seen as strategic, since it creates goodwill, it helps to understand the interests of the forum, and to anticipate its reactions. Similarly, Welsh public sector managers have reported that collaborative and supportive relationship with their regulators contributes to the improvement of the achievement of their professional mission (Andrews et al. 2008). Thus, transforming the character of the accountability relationship from that of control and oversight to one of trust and collaboration, eases the communication between the actor and the forum, and gives the actor more freedom and power to pursue its interests (Busuioc and Lodge 2017; Romzek, LeRoux, and Blackmar 2012). Following the strategic considerations of public sector actors, we would expect that they give priority to the demands of the forums with which they have already had positive experience, as that helps to build and maintain positive accountability relationships.

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

Study Design and Data
Experimental Design
This study was conducted using a conjoint experimental design. Conjoint designs allow for causal testing of the effects of multiple treatment components simultaneously and enable scholars to estimate the relative effect of each of them (Green and Rao 1971). In this type of experimental study, respondents choose between and/or evaluate a set of alternatives, each having a varied set of attributes (Hainmueller, Hopkins, and Yamamoto 2014). The method of conjoint analysis is still rare but it is being increasingly used in the fields of political science and public administration in recent years, due to its many advantages, such as its power to simultaneously investigate several causal claims, as well as due to its simplicity (for example Hollibaugh, Miles, and Newswander 2020; Jilke and Tummers 2018). Conjoint studies limit the effects of social desirability bias, since they allow respondents to justify their decisions using multiple different reasons (Hainmueller and Hopkins 2015). Furthermore, the choices actors make in the real world, such as choices of public sector actors in multiple accountabilities environments, are often made as a result of trade-offs and interplay of different characteristics of the decision-making environment. Through their design, conjoint experiments account for these trade-offs, creating results with higher degrees of realism (Hainmueller, Hopkins, and Yamamoto 2014). Finally, the use of conjoint experiments in this context helps diversify the methodological approaches used to study public accountability (Yang 2012).

In this study, respondents were asked to decide which stakeholder to prioritize from a given set of two profiles (the scenario is provided in appendix). Each stakeholder was described using four attributes, which capture the variables of interest: likelihood of consequences for non-compliance, type of consequence for non-compliance, stakeholder expertise, and previous experience with stakeholder. The overview of the attributes and their levels is provided in Table 1. The two stakeholder profiles were randomly generated from the list of available levels of each attribute.

Table 1 Overview of the Attributes and Their Levels

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

Table 2 Overview of Subject Characteristics

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

Data

We performed two rounds of data collection. First, the experiment was run on a sample of 600 civil servants in the United Kingdom, recruited through the online subject pool Prolific.ac (Palan and Schitter 2018) in July 2019.[1] Each participant was asked to make four prioritization decisions, and we thus obtained 2,345[2] observations of prioritization choices. A second sample was collected in October 2019 on 603 civil servants in the Netherlands, working in four executive agencies providing direct financial benefits or grants to citizens. These organizations provide income support (UWV), tax allowances (Belastingdienst-Toeslagen), student loans (DUO), and health care benefits (CAK). The selection and recruitment of participants was done within the organizations themselves. All organizations were instructed to recruit participants who enjoy a certain degree of autonomy in their work, which makes them a subject to accountability pressures. Here again, respondents were asked to make four prioritization decisions, and we obtained 2,292 observations of prioritization choices. Table 2 provides an overview of the characteristics of the two samples.

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

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

First of all, bureaucracy in the United Kingdom has "traditionally developed under the spell of the political domain" (Kuhlmann and Wollmann 2019, 22). Thus, the civil service in the United Kingdom is more tightly controlled and centralized than in other European countries (Hill and Varone 2014, 87). In contrast, the Dutch public sector is traditionally fragmented (Andeweg and Irwin 2014) and the majority of civil servants, including those surveyed by us, work in agencies at arms' length of central government (Van Thiel and Yesilkagit 2011). This distinction is relevant for their experiences with multiple accountabilities. Specifically, the accountability cross-pressures, stemming from various accountability forums, our independent variable, would be experienced more strongly by civil servants in fragmented systems, such as in the Netherlands, than by civil servants in more concentrated and politically controlled systems, as in the United Kingdom.

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

External and Ecological Validity Considerations

Our scenario presents an abstract accountability forum prioritization task, in which the stakeholders are only described with the factors of our interest, while all other context is removed. This approach offers a number of advantages. First, our general scenario allows to investigate various types of civil servants in both

countries simultaneously. While multiple accountability dilemmas are likely to present themselves somewhat differently for public sector managers and for street level bureaucrats, our scenario allows the civil servants to associate it with the dilemmas they specifically face in their own workplace. Secondly, from a methodological point of view, providing more context is likely to introduce noise into the experiment. For instance, a specific identity of a forum is likely to elicit experiences and attitudes toward those specific stakeholders, which we would not be able to account for in our analysis.

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

Analysis

Using the package "cregg" in R (Leeper 2018), we estimate the marginal mean (MM) for each attribute level in our study. The marginal mean reflects the probability that a certain stakeholder is prioritized given the presence of a particular attribute level (Leeper,

Hobolt, and Tilley 2019). It displays the effect of an attribute level with reference to a null effect, or the probability that the stakeholder profile with the particular attribute is chosen completely randomly.

Results

We present our results in two steps. We first present the main results based on the pooled sample, after which we examine the cross-country variation. Additional analyses examining organizational patterns in the Dutch sample and the influence on personal and professional characteristics on prioritization choices are provided in an appendix.

Overall Results

The results of the conjoint analysis on the pooled sample are presented in Figure 1, where the y-axis presents the attribute levels, while the x-axis presents the marginal means. The vertical line at marginal mean .5 presents equal likelihood of being prioritized given the specific attribute level. Thus, estimates to the right of the vertical line present higher than equal likelihoods, while estimates to the left of the vertical line present lower than equal likelihoods that the stakeholder with the given attribute level will be prioritized. The closer the marginal mean estimate is to the vertical line, the smaller the effect of the attribute level on the prioritization choice. The horizontal lines running through the marginal means estimates for each attribute level, represent the 95% confidence intervals.
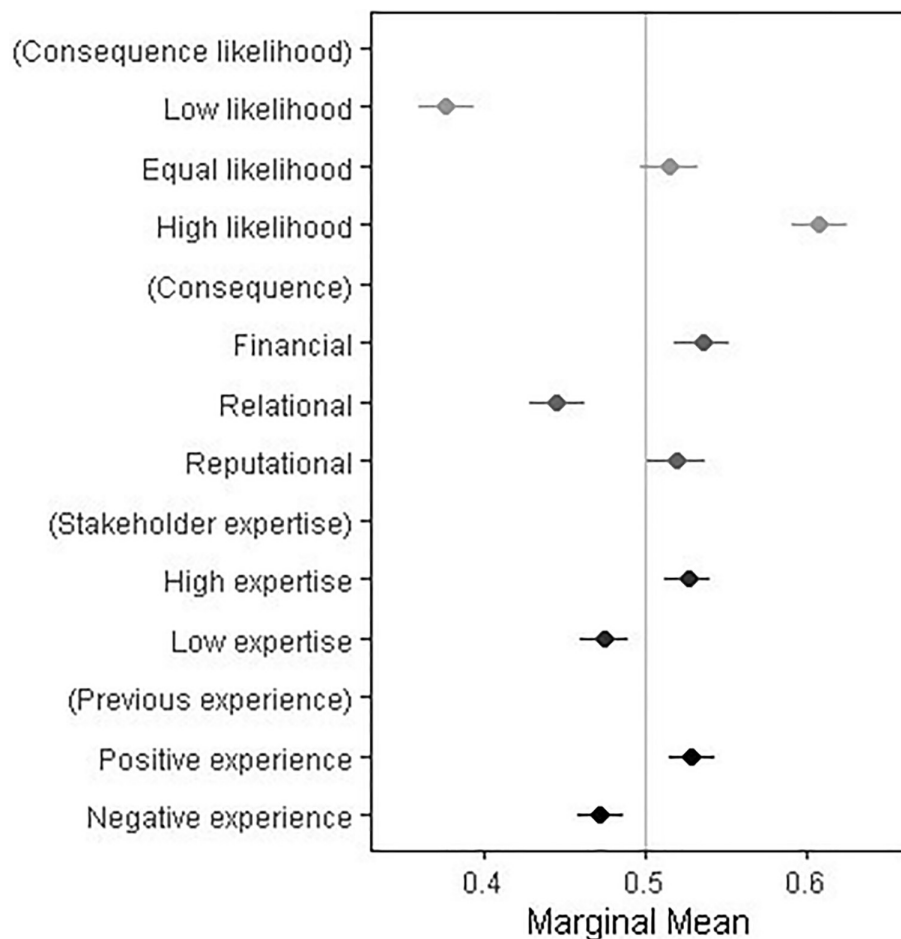
Note: Figure displays estimated marginal means with confidence intervals set at 95 percent

Figure 1   Accountability Forum Prioritization—Pooled Sample (N = 4,637).

The results displayed in Figure 1 indicates that the likelihood of sanction is by far the strongest driver of prioritization choices. Specifically, a high likelihood that the stakeholder will impose a consequence on the public sector actor who does not comply with its demand significantly increases the likelihood of prioritization (MM = .61, SE = .01), while a low likelihood significantly decreases it (MM = .38, SE = .01). This is in line with our expectations outlined in H1.

Additionally, we also find that the other elements have some impact on prioritization choices, yet to a much lesser degree. The type of sanction the stakeholder threatens to impose in the case of non-compliance with demands also plays a role. Our analysis indicates that stakeholders who threaten to impose financial (MM = .53, SE = .01) or reputational sanctions (MM = .52, SE = .01) are significantly more likely to be prioritized than stakeholders who threaten to impose relational sanctions (MM = .45, SE = .01). The results partially support H2, in that while we find support that formal financial sanctions will be perceived as more important than informal ones, we do not find a statistically significant difference between the effect of the financial and reputational sanctions.

Stakeholder expertise also plays a role in prioritization choices. In line with H3, we do find that stakeholders with high expertise are more likely to be prioritized (MM = .53, SE = .01) over others. Finally, previous experiences of dealing with the stakeholder do influence the prioritization choices of civil servants in that positive previous experience is given priority (MM = .53, SE = .01). This finding is in line with our expectations in H4 (a table with MM estimates is provided in the appendix).

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

Country Comparison

To assess the robustness of the findings in the pooled sample analysis, we evaluate whether they replicate in both country samples independently. Because both samples differ in their composition, specifically, the U.K. sample contains more managers and highly educated respondents than the Dutch one (see Table 2), we performed propensity score matching and extracted subsamples to ensure samples are comparable in terms of subject characteristics (Rosenbaum and Rubin 1985). The propensity score matching was conducted using the package "MatchIt" in R (Ho et al. 2011), with the method nearest neighbor, discarding unmatched observations from both samples, using a caliper of .2. Caliper is a measure of how "different" subsamples are allowed to be and a caliper of .2 has been found to be optimal (Austin 2011). The matching procedure produced subsamples consisting of 312 respondents from each country which are not statistically different on any of the personal and professional characteristics.[3] The subject characteristics post-matching are provided in the appendix. As a robustness check, we performed the analysis with both matched and full samples and this yields no substantively different results.[4]

The results of the analysis of the matched samples are presented in Figure 2. An omnibus F-test, comparing two nested models, one estimating only the effect of the attributes, and one including a country interaction (Leeper, Hobolt, and Tilley 2019), provides evidence of similarities but also some smaller, yet significant differences in the patterns of prioritization choices of civil servants in the two countries ($F[7; 4,866] = 6.52$, $p < .01$). As Figure 2 shows, the prioritization choices of civil servants in both countries are influenced by the threat of sanction in the same way: both groups are strongly driven by the likelihood of sanction. However, this effect is significantly stronger for Dutch civil servants (high likelihood of sanction MM = .65, SE = .02) than for U.K. civil servants (high likelihood of sanction MM = .58, SE = .02). Furthermore, when it comes to the effects of stakeholder expertise and previous experience with the stakeholder, we observe some differences in the responses of the civil servants in the two countries. Specifically, while U.K. civil servants do prioritize stakeholders with higher expertise (MM = .55, SE = .01), this is not the case for Dutch civil servants (MM = .49, SE = .01). U.K. civil servants also place importance on the previous experience with the stakeholder, and thus give priority to stakeholders with whom they have had positive previous experience (MM = .55, SE = .01), although the same tendency can be observed among the Dutch civil servants, the effect is much weaker (MM = .52, SE = .01, respectively).

Discussion

Our findings provide support for all four hypotheses, although not to the same degree. The credible threat of sanction increases the likelihood that a forum will receive priority (H1). Furthermore, formal financial sanctions are generally prioritized over informal ones (H2). H3 was also supported since forums with high expertise are favored over those with low expertise. Finally, positive prior experiences with a forum increases the likelihood of prioritizing that forum (H4).

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

In contrast to the consistency of the effects of sanctions on stakeholder prioritization, we found that the effects of stakeholder
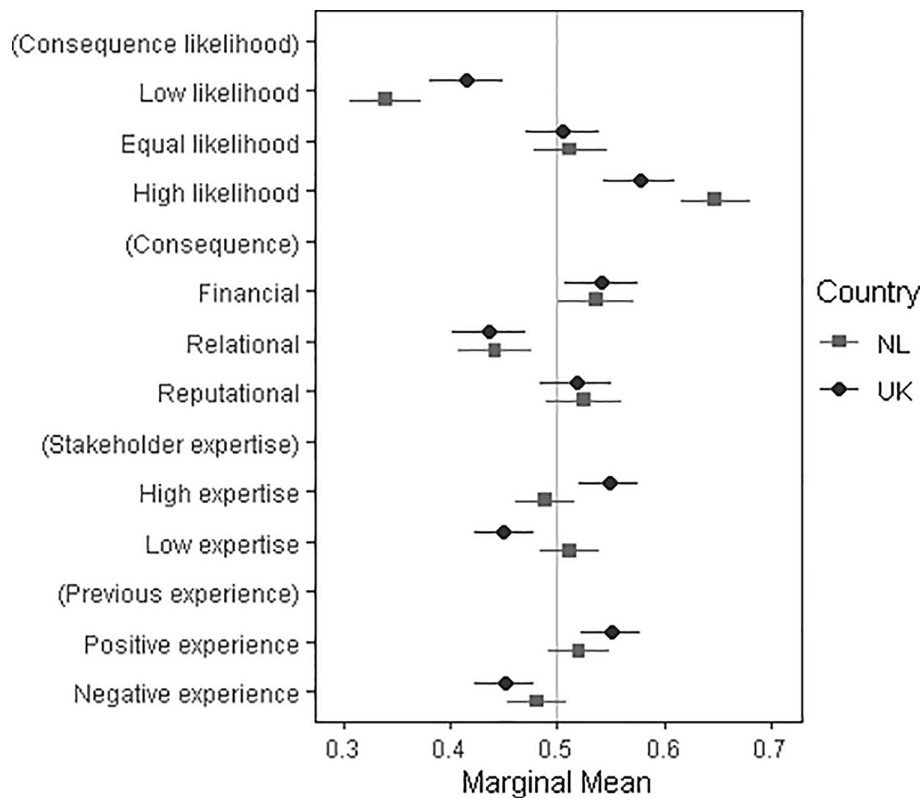
Note: Figure displays estimated marginal means with confidence intervals set at 95 percent

Figure 2   Accountability Forum Prioritization Per Country—Matched Sample (N = 2,441).

expertise and the relationship with the stakeholder vary somewhat across contexts. Thus, while stakeholder expertise led to significant stakeholder prioritization in the United Kingdom, this effect was not observed in the Netherlands. Furthermore, while in both countries there was a tendency to prioritize stakeholders with whom there was positive previous experience, this effect is much weaker in the Netherlands. Different institutional arrangements as well as values could be the drivers of these differences. The greater fragmentation and stronger commitment to equal treatment of the Dutch civil service, as compared to the United Kingdom, could be a part of the explanation (Dahlström et al. 2015). These tentative explanations, however, should be subjected to further investigation. They suggest, overall, that accountability as *management of expectations* is relevant yet less universal and more context-dependent than accountability as *answerability*.

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

they nevertheless have secondary importance in account-holding and enforcement.

Second, the prioritization choices of public sector decision-makers speak directly to the relative powers of the forums. Evidently, the ones that are able to pose the most credible threat, especially of financial or reputational nature, will be able to exert significant impact on the work of the public sector actor. This might give the upper hand to forums with more aggressive enforcement styles and greater resources to inflict damage. Furthermore, the prioritization of forums with which the decision-maker has had good working experiences signifies that building a good rapport with the accountee could also give the forum a favored position, although less effectively than with the threat to impose a sanction. This type of collaborative relationship between the forum and the actor could be seen as preferable to one of control, as it is less antagonistic, reduces oversight costs (Romzek, LeRoux, and Blackmar 2012), and increases the willingness of the public sector actor to comply with the forum's demands (Andrews et al. 2008). Yet, collaborative relationships could turn into over-comfortable ones, resulting into impunity or regulatory capture (Braun 2012).

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

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

Second, our experimental scenario presents a stylized task in which the decision-making context is removed. While this approach eases the application of the research tool to various contexts, and thereby facilitates comparability of groups of respondents, it does not account for the nuances of the decision-making context in which a public sector actor is situated. Specific organizational-level dynamics, for example, cannot be explored in much detail. Our results could be, thus, treated as a baseline, to which the decision-making in specific multiple accountability environments can be compared. A case in point is a recent study by Hollibaugh, Miles, and Newswander (2020) which shows that in conditions where the forum demands are seen as potentially harmful or unethical, hierarchy and potential sanctions play an insignificant role.

Third, while we use an original dataset, containing two national samples, our data has some limitations. Specifically, our samples are not representative of an underlying population[5] and they have been collected using different sampling methods in different periods of the year. The differences in the sampling may translate in dissimilarities in sample composition. We accounted for this using the method of propensity score matching (Ho et al. 2011; Rosenbaum and Rubin 1985). This matching method, however, makes the samples comparable on the basis of measured participant characteristics and is unable to account for unmeasured characteristics that could affect decision-making behavior. The domain of work, the particular stakeholders the respondents are regularly confronted with, the nature of the stakeholder demands, thus again, the context of the decision, could influence prioritization choices in multiple accountability dilemmas.

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

Conclusion

This study investigated the way public sector actors assign priority to accountability forums in multiple accountabilities dilemmas. The necessity to simultaneously give account to multiple stakeholders has been noted as a universal, and very

often problematic phenomenon in the public sector (Hupe and Hill 2007; Koppell 2005; Thomann, Hupe, and Sager 2018). The way in which the decision-making behavior of public sector actors is shaped as a result of this phenomenon is, however, considerably less understood (Aleksovska, Schillemans, and Grimmelikhuijsen 2019; Yang 2012). Accountability has often been defined either as "answerability" or as "managing expectations". Our study now suggests that these two processes are both relevant yet not in the same way. Answerability, most notably pre-empting sanctions (financial or reputational), seems to be a universal process preceding other considerations. Managing expectations plays second fiddle, yet seems to be important in explaining more fine-grained contextual differences.

Notes

1 The respondents were compensated 0.60 pounds (0.75 U.S. dollars).
2 Observations where by chance the respondents were asked to choose from two identical stakeholders were removed from the sample.
3 The matching was performed based on all subject characteristics and the extent to which the respondents recognized the dilemma provided in the scenario.
4 The effect of financial sanction is an exception: it is greater in the U.K. sample than in the Dutch sample, but post-matching this difference disappears.
5 Additionally, we are unable to test for non-response bias.

References

Aberbach, Joel D., and Bert A. Rockman. 1997. Bureaucracy: Control, Responsiveness, Performance. In *Designs for Democratic Stability: Studies in Viable Constitutionalism*, edited by Abdo I. Baaklini and Helen Desfosses, 73–94. New York: Sharpe.

Acar, Muhittin, Chao Guo, and Kaifeng Yang. 2008. Accountability When Hierarchical Authority Is Absent: Views From Public–Private Partnership Practitioners. *The American Review of Public Administration* 38(1): 3–23.

Aleksovska, Marija, Thomas Schillemans, and Stephan Grimmelikhuijsen. 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.

Andeweg, Rudy B., and Galen A. Irwin. 2014. *Governance and Politics of The Netherlands*. Macmillan International Higher Education.

Andrews, Rhys, George A. Boyne, Jennifer Law, and Richard M. Walker. 2008. Organizational Strategy, External Regulation and Public Service Performance. *Public Administration* 86(1): 185–203.

Ashworth, Rachel, George A. Boyne, and Richard M. Walker. 2002. Regulatory Problems in the Public Sector: Theories and Cases. *Policy & Politics* 30(2): 195–211.

Austin, Peter C. 2011. Optimal Caliper Widths for Propensity-Score Matching when Estimating Differences in Means and Differences in Proportions in Observational Studies. *Pharmaceutical Statistics* 10(2): 150–61.

Bach, Tobias, Sandra van Thiel, Gerhard Hammerschmid, and Reto Steiner. 2017. Administrative Tradition and Management Reforms: A Comparison of Agency Chief Executive Accountability in Four Continental Rechtsstaat Countries. *Public Management Review* 19(6): 765–84.

Benjamin, Lehn M., and Paul L. Posner. 2018. Tax Expenditures and Accountability: The Case of the Ambivalent Principals. *Journal of Public Administration Research and Theory* 28(4): 569–82.

Bovens, Mark. 2007. Analysing and Assessing Accountability: A Conceptual Framework. *European Law Journal* 13(4): 447–68.

Boyne, George, Patricia Day, and Richard Walker. 2002. The Evaluation of Public Service Inspection: A Theoretical Framework. *Urban Studies* 39(7): 1197–212.

Braithwaite, John. 1997. On Speaking Softly and Carrying Big Sticks: Neglected Dimensions of a Republication Separation of Powers. *The University of Toronto Law Journal* 47(3): 305–61.

Braun, Caelesta. 2012. The Captive or the Broker? Explaining Public Agency–Interest Group Interactions. *Governance* 25(2): 291–314.

Busuioc, Madalina. 2009. Accountability, Control and Independence: The Case of European Agencies. *European Law Journal* 15(5): 599–615.

Busuioc, Madalina, and Martin Lodge. 2017. Reputation and Accountability Relationships: Managing Accountability Expectations through Reputation. *Public Administration Review* 77(1): 91–100.

Dahlström, Carl, Jan Teorell, Stefan Dahlberg, Felix Hartmann, and Annika Lindberg. 2015. *The QoG Expert Survey II Report*.

Ewert, Benjamin. 2018. Focusing on Quality Care Rather than 'Checking Boxes': How to Exit the Labyrinth of Multiple Accountabilities in Hybrid Healthcare Arrangements. *Public Administration*: 1–17.

Flinders, Matthew, and Katherine Tonkiss. 2016. From 'Poor Parenting' to Micro-Management: Coalition Governance and the Sponsorship of Arm's-Length Bodies in the United Kingdom, 2010–13. *International Review of Administrative Sciences* 82(3): 490–515.

Gailmard, Sean. 2014. Accountability and Principal-Agent Models. In *The Oxford Handbook of Public Accountability*, edited by Mark Bovens, Robert E. Goodin, and Thomas Schillemans, 90–106. Oxford: Oxford University Press.

Green, Paul E., and Vithala R. Rao. 1971. Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research* 8(3): 355–63.

Greiling, Dorathea, and Katharina Spraul. 2010. Accountability and the Challenges of Information Disclosure. *Public Administration Quarterly* 34(3): 338–77.

Hainmueller, J., and Hopkins D. J. 2015. The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes Toward Immigrants. *American Journal of Political Science* 59(3): 529–548.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices Via Stated Preference Experiments. *Political Analysis* 22(1): 1–30.

Hill, Carolyn J., and Laurence E. Lynn. 2004. Is Hierarchical Governance in Decline? Evidence from Empirical Research. *Journal of Public Administration Research and Theory* 15(2): 173–95.

Hill, Michael, and Frederic Varone. 2014. *The Public Policy Process*. London: Routledge.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 42(8): 1–28.

Hollibaugh, G.E., Jr., M.R. Miles, and C.B. Newswander. 2020. Why Public Employees Rebel: Guerrilla Government in the Public Sector. *Public Administration Review* 80(1): 64–74.

Hood, Christopher, Oliver James, Colin Scott, George W. Jones, and Tony Travers. 1999. *Regulation inside Government: Waste Watchers, Quality Police, and Sleaze-Busters*. Oxford: Oxford University Press.

Hupe, Peter, and Michael Hill. 2007. Street-Level Bureaucracy and Public Accountability. *Public Administration* 85(2): 79–299.

Jilke, Sebastian, and Lars Tummers. 2018. Which Clients Are Deserving of Help? A Theoretical Model and Experimental Test. *Journal of Public Administration Research and Theory* 28(2): 226–38.

Koppell, Jonathan G.S. 2005. Pathologies of Accountability: ICANN and the Challenge of "Multiple Accountabilities Disorder". *Public Administration Review* 65(1): 94–108.

Kuhlmann, Sabine, and Hellmut Wollmann. 2019. *Introduction to Comparative Public Administration: Administrative Systems and Reforms in Europe*. Cheltenham: Edward Elgar Publishing.

Leeper, Thomas J. 2018. Cregg: Simple Conjoint Analyses and Visualization. R Package Version 0.2.1.

Leeper, Thomas J., Sara B. Hobolt, and James Tilley. 2019. Measuring Subgroup Preferences in Conjoint Experiments. *Political Analysis*: 1–15.

Lerner, Jennifer S., and Philip E. Tetlock. 1999. Accounting for the Effects of Accountability. *Psychological Bulletin* 125(2): 255–75.

Lieberherr, Eva, and Eva Thomann. 2019. Street-Level Bureaucracy Research and Accountability beyond Hierarchy. In *Research Handbook on Street-Level Bureaucracy*, edited by Peter Hupe, 223–39. Cheltenham: Edward Elgar Publishing.

Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.

O'Loughlin, Michael G. 1990. What Is Bureaucratic Accountability and How Can We Measure it? *Administration and Society* 22(3): 275–302.

Overman, Sjors, Thomas Schillemans, and Stephan Grimmelikhuijsen. 2020. A Validated Measurement for Felt Relational Accountability in the Public Sector: Gauging the Account holder's Legitimacy and Expertise. *Public Management Review*: 1–20.

Palan, Stefan, and Christian Schitter. 2018. Prolific.ac—A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance* 17: 22–7.

Romzek, Barbara S., and Melvin J. Dubnick. 1987. Accountability in the Public Sector: Lessons from the Challenger Tragedy. *Public Administration Review* 47(3): 227–38.

Romzek, Barbara S., Kelly LeRoux, and Jeannette M. Blackmar. 2012. A Preliminary Theory of Informal Accountability among Network Organizational Actors. *Public Administration Review* 72(3): 442–53.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician* 39(1): 33–8.

Schillemans, Thomas. 2015. Managing Public Accountability: How Public Managers Manage Public Accountability. *International Journal of Public Administration* 38(6): 433–41.

———. 2016. Calibrating Public Sector Accountability: Translating Experimental Findings to Public Sector Accountability. *Public Management Review* 18(9): 1400–20.

Schillemans, Thomas, and Mark Bovens. 2011. The Challenge of Multiple Accountability. In *Accountable Governance: Problems and Promises*, edited by J. Melvin and H. George Frederickson Dubnick, 3–21. London: Routledge.

Schillemans, Thomas, and Madalina Busuioc. 2015. Predicting Public Sector Accountability: From Agency Drift to Forum Drift. *Journal of Public Administration Research and Theory* 25(1): 191–215.

Schwabenland, Christina, and Alison Hirst. 2018. Hybrid Accountabilities and Managerial Agency in the Third Sector. *Public Administration*: 1–15.

Shortell, Stephen M., Teresa M. Waters, Kenneth W.B. Clarke, and Peter P. Budetti. 1998. Physicians as Double Agents: Maintaining Trust in an era of Multiple Accountabilities. *JAMA* 280(12): 1102–8.

Sinclair, Amanda. 1995. The Chameleon of Accountability: Forms and Discourses. *Accounting, Organizations and Society* 20(2–3): 219–37.

Tetlock, Philip E. 1992. The Impact of Accountability on Judgment and Choice: Toward a Social Contingency Model. *Advances in Experimental Social Psychology* 25(3): 331–76.

Thomann, Eva, Peter Hupe, and Fritz Sager. 2018. Serving Many Masters: Public Accountability in Private Policy Implementation. *Governance* 31(2): 299–319.

Van Thiel, Sandra, and Kutsal Yesilkagit. 2011. Good Neighbours or Distant Friends? Trust between Dutch Ministries and Their Executive Agencies. *Public Management Review* 13(6): 783–802.

Willems, Tom, and Wouter van Dooren. 2012. Coming to Terms with Accountability: Combining Multiple Forums and Functions. *Public Management Review* 14(7): 1011–36.

Yang, Kaifeng. 2012. Further Understanding Accountability in Public Organizations: Actionable Knowledge and the Structure–Agency Duality. *Administration and Society* 44(3): 255–84.

Appendix A: Experimental Scenario

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

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

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

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

(Example profile choice)

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

Appendix B: Sampling Procedure for the U.K. Sample

We sampled using two presets in Prolific.ac: country of residence and employer type. We chose United Kingdom as a country of residence and not nationality since nationality is not a pre-requisite for employment in the public sector (is most cases), and because U.K. nationals can potentially work in the public sector of other countries. We aimed to capture respondents who work in the U.K. public sector, and thus country of residence is in our opinion more appropriate than nationality for that goal. For employer type we selected local, state and federal government employee. There were 1,125 respondents in prolific's pool that have been active in the last 90 days and matched these characteristics at the time of our sampling. In addition we added two yes/no questions at the start of our survey which asked the participants 1) Do you currently reside in the United Kingdom? 2) Are you employed by a governmental or a public sector organization?. If they answered negatively to at least one of the questions, they were not permitted to proceed with the survey.

Appendix C: Overview of the Surveyed Organizations and Samples in the Netherlands

	Number of Employees	Parent Ministry	Completed Responses	Response Rate
Belastingdienst Toeslagen (BD-TL)	~28,000	Ministry of finance	111	43.5 percent
Centraal Administratie Kantoor (CAK)	~1,250	Ministry of health, welfare and sport	95	63.3 percent
Dienst Uitvoering Onderwijs (DUO)	~3,000	Ministry of education, culture and science	181	38.1 percent
Uitvoeringsinstituut Werknemersverzekeringen (UWV)	~17,000	Ministry of social affairs and employment	216	79.4 percent
			126 Email	.5 percent
			90 Intranet	

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

Appendix D: Validity Considerations

Do you recognize the situation provided in this study in your work?

	U.K.	NL				
		BD-TL	CAK	DUO	UWV	All
I deal with similar situations very often	64 (10.7 percent)	8 (7.2 percent)	25 (26.3 percent)	13 (7.2 percent)	14 (6.5 percent)	60 (10.0 percent)
I sometimes deal with similar situations	202 (33.7 percent)	51 (45.9 percent)	40 (42.1 percent)	85 (47.0 percent)	112 (51.9 percent)	288 (47.8 percent)
I recognize these situations in the work of my colleagues, but not in my own work	160 (26.7 percent)	20 (18.0 percent)	24 (25.3 percent)	33 (18.2 percent)	43 (19.9 percent)	120 (19.9 percent)
Such situations are very rare in my workplace	117 (19.5 percent)	9 (8.1 percent)	2 (2.1 percent)	24 (13.3 percent)	26 (12.0 percent)	61 (10.1 percent)
I have never heard of, or encountered such a situation before	57 (9.5 percent)	19 (17.1 percent)	2 (2.1 percent)	15 (8,3 percent)	13 (6.0 percent)	49 (8.1 percent)
(No answer)	0 (.0 percent)	4 (3.6 percent)	2 (2.1 percent)	11 (6.1 percent)	8 (3.7 percent)	25 (4.1 percent)
N	600	111	95	181	216	603

Appendix E: Overview of Sample Characteristics Post Matching

	U.K. Sample		Dutch Sample	
	N	Percent	N	Percent
Gender				
Male	83	26.6 percent	92	29.5 percent
Female	229	73.4 percent	220	70.5 percent
Education				
Secondary school or lower	17	5.5 percent	13	4.2 percent
Post-secondary school qualifications	85	27.2 percent	85	27.2 percent
Bachelor's degree or higher	210	67.3 percent	214	68.6 percent
Manager				
Yes	34	10.9 percent	28	9.0 percent
No	278	89.1 percent	284	91.0 percent
Supervisor				
Yes	82	26.3 percent	76	24.4 percent
No	230	73.7 percent	236	75.6 percent
Client interaction				
Very often	161	51.6 percent	151	48.4 percent
Often	59	18.9 percent	74	23.7 percent
Sometimes	42	13.5 percent	43	13.8 percent
Rarely	26	8.3 percent	17	5.4 percent

|	U.K. Sample		Dutch Sample	
	N	Percent	N	Percent
Never	24	7.7 percent	27	8.7 percent
	Mean (SD)	Min/Max	Mean (SD)	Min/Max
Age	43.0 (10.7)	20/71	43.5 (11.0)	20/67
Work tenure	11.0 (8.8)	.2/41	11.7 (9.4)	0/39
Total N	312		312	

Appendix F: Organizational Differences

Surveyed Organizations

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

Analysis

Do the general patterns observed for the Dutch civil servants replicate on an organizational level? The analysis of the prioritization decisions for the respondents from the four surveyed Dutch executive agencies is presented in the figure below. An omnibus F-test points out to some smaller yet significant organizational differences ($F[21; 4,556] = 3.28$, $p < .01$). As it can be observed from the figure, while all organizations respond in a similar way to the threat of sanction—prioritize when the threat is high and de-prioritize when the threat is low—their sensitivity to this threat differs. Thus, we see CAK (*Centraal Administratie Kantoor)* civil servants, for example, a much more risk averse than the civil servants from the other organizations, since they assign higher priority to stakeholders even when the threat of sanction is 50–50. When it comes to the effects of the different types of sanctions, we observe the general pattern of prioritizing stakeholders who threaten financial and reputational sanction more than ones that threaten a relational sanction. Here again, CAK civil servants stand out somewhat due to the importance they give to the reputational sanction. In terms of stakeholder expertise, we observe a consistent absence of effect. Finally, regarding the effect of previous experience on prioritization decisions, we observe some variation: DUO employees give priority to stakeholders with whom they have had positive previous experience, CAK has the tendency to do the opposite, while this does not bear any importance for BD-TL and UWV employees.
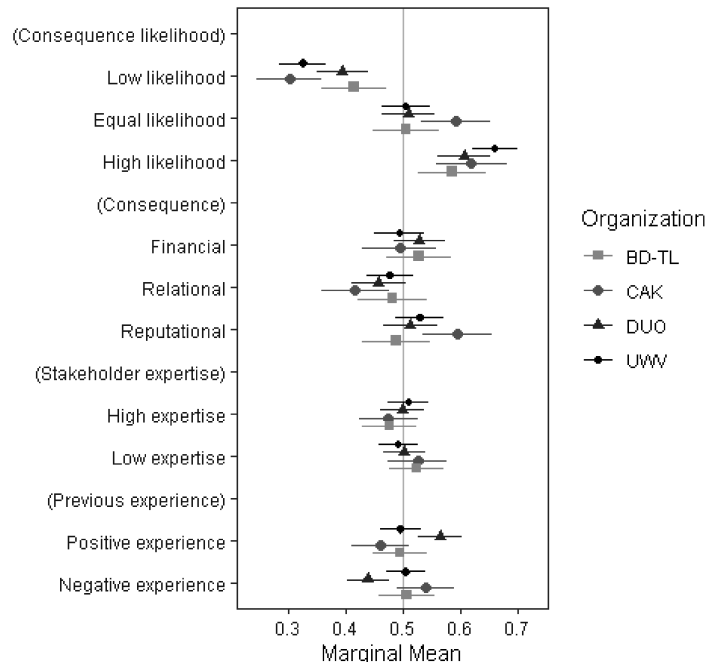

The results again indicate that the effect of sanctions is consistent, although CAK appears to be an outlier. The CAK is an executive agency that carries out various social security regulations for the government and is generally less well-known among the general public than the other three agencies.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

regarding stakeholder prioritization. Our study, however, is not able to probe further into the sources of these organizational and contextual effects.

In order to assess whether the organizational differences observed here have a strong bearing on the overall results, we conduct a sensitivity analysis whereby we systematically exclude one the of the organizations from the overall sample one-by-one, and compare the reduced samples to each other and to the overall sample. In this way, we are able to assess whether a single-organizational sample drives the results significantly into a particular direction. The results of this analysis are presented in the figure below. An omnibus test ($F[28; 18,301] = .80$, $p = .76$) indicates that there is no significant difference in the patterns displayed by the five samples.

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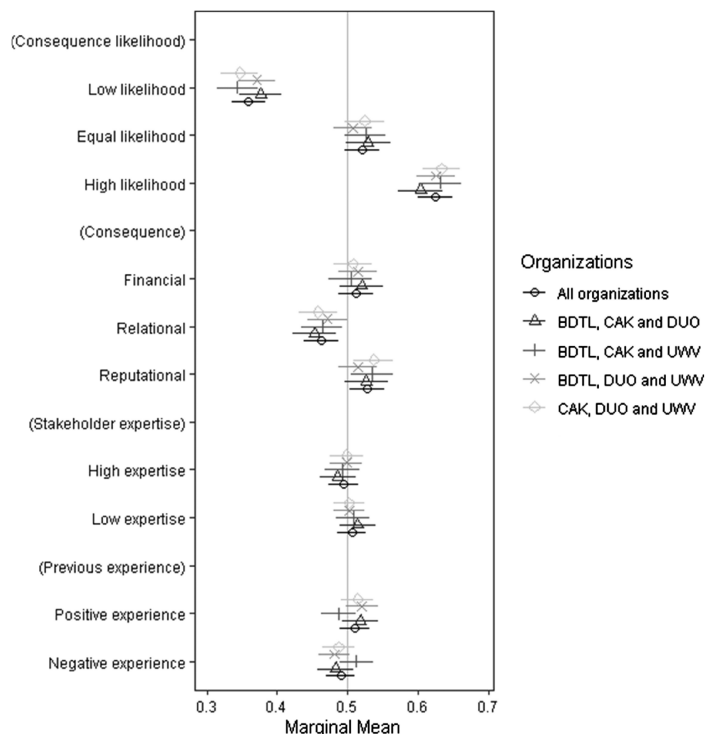
Appendix G: Detailed Results
Marginal means (standard errors in brackets).

	All	U.K.	NL	U.K.-Matched	NL-Matched	BD-TL	CAK	DUO	UWV
Consequence likelihood									
Low likelihood	.38 (.01)	.39 (.01)	.36 (.01)	.42 (.02)	.34 (.02)	.41 (.03)	.30 (.03)	.39 (.02)	.33 (.02)
Equal likelihood	.51 (.01)	.51 (.01)	.52 (.01)	.50 (.02)	.51 (.02)	.51 (.03)	.59 (.03)	.51 (.02)	.51 (.02)
High likelihood	.61 (.01)	.59 (.01)	.62 (.01)	.58 (.02)	.65 (.02)	.58 (.03)	.62 (.03)	.61 (.02)	.66 (.02)
Consequence									
Financial	.53 (.01)	.56 (.01)	.51 (.01)	.54 (.02)	.54 (.02)	.53 (.03)	.49 (.03)	.53 (.02)	.49 (.02)
Relational	.45 (.01)	.43 (.01)	.46 (.01)	.44 (.02)	.44 (.02)	.48 (.03)	.42 (.03)	.46 (.02)	.48 (.02)
Reputational	.52 (.01)	.51 (.01)	.53 (.01)	.52 (.02)	.52 (.02)	.49 (.03)	.59 (.03)	.51 (.02)	.53 (.02)
Stakeholder expertise									
High expertise	.53 (.01)	.56 (.01)	.49 (.01)	.55 (.01)	.49 (.01)	.48 (.02)	.47 (.03)	.50 (.02)	.51 (.02)
Low expertise	.47 (.01)	.44 (.01)	.51 (.01)	.45 (.01)	.51 (.01)	.52 (.02)	.53 (.03)	.50 (.02)	.49 (.02)
Previous experience									
Positive experience	.53 (.01)	.55 (.01)	.51 (.01)	.55 (.01)	.52 (.01)	.49 (.01)	.46 (.03)	.56 (.02)	.50 (.02)
Negative experience	.47 (.01)	.45 (.01)	.49 (.01)	.45 (.01)	.48 (.01)	.51 (.02)	.54 (.03)	.44 (.02)	.50 (.02)

Appendix H: Additional Analysis

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

	All	U.K.	NL	BD-TL	CAK	DUO	UWV
Personal characteristics							
Gender	$F(7; 9,042) = 2.81$, $p = .01$	$F(7; 4,660) = .38$, $p = .92$	$F(7; 4,368) = 1.40$, $p = .20$	$F(7; 776) = .56$, $p = .79$	$F(7; 714) = 1.57$, $p = .14$	$F(7; 1,272) = 1.57$, $p = .14$	$F(7; 1,564) = .41$, $p = .89$
Education	$F(14; 9,137) = 2.84$, $p < .01$	$F(14; 4,669) = 2.42$, $p < .01$	$F(14; 4,447) = .94$, $p = .52$	$F(14; 793) = .67$, $p = .80$	$F(14; 713) = 1.53$, $p = .09$	$F(14; 1,297) = 2.05$, $p = .01$	$F(14; 1,581) = 1.38$, $p = .15$
Age	$F(14; 8,783) = 1.68$, $p = .05$	$F(14; 4,613) = 1.62$, $p = .07$	$F(14; 4,149) = 1.27$, $p = .21$	$F(14; 755) = 1.32$, $p = .19$	$F(14; 681) = .74$, $p = .74$	$F(14; 1,181) = 1.96$, $p = .02$	$F(14; 1,469) = .89$, $p = .57$
Professional characteristics							
Manager	$F(7; 9,152) = 2.55$, $p = .01$	$F(7; 4,668) = 4.96$, $p < .01$	$F(7; 4,470) = .92$, $p = .49$	$F(7; 808) = 2.19$, $p = .03$	$F(7; 720) = .63$, $p = .73$	$F(7; 1,312) = .34$, $p = .94$	$F(7; 1,588) = .75$, $p = .63$
Supervisor	$F(7; 9,144) = 1.07$, $p = .38$	$F(7; 4,676) = 2.02$, $p = .03$	$F(7; 4,454) = 1.46$, $p = .18$	$F(7; 800) = 1.29$, $p = .25$	$F(7; 720) = .17$, $p = .99$	$F(7; 1,312) = 1.41$, $p = .19$	$F(7; 1,580) = 1.04$, $p = .40$
Tenure	$F(14; 8,799) = 2.03$, $p = .01$	$F(14; 4,621) = 2.26$, $p < .01$	$F(14; 4,157) = 1.33$, $p = .18$	$F(14; 745) = .94$, $p = .52$	$F(14; 657) = 1.29$, $p = .21$	$F(14; 1,179) = 1.82$, $p = .03$	$F(14; 1,513) = .99$, $p = .45$
Client interaction	$F(14; 9,153) = 1.12$, $p = .34$	$F(14; 4,669) = .91$, $p = .54$	$F(14; 4,463) = .61$, $p = .86$	$F(14; 801) = .50$, $p = .93$	$F(14; 713) = 1.13$, $p = .33$	$F(14; 1,305) = .75$, $p = .72$	$F(14; 1,581) = 1.30$, $p = .20$

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