A reputational perspective on structural reforms:

How media reputations are related to the structural reform likelihood of public agencies

Jan Boon, University of Antwerp

Jan Wynen, University of Antwerp

Koen Verhoest, University of Antwerp

Walter Daelemans, University of Antwerp

Introduction

In this study we theorize and test how the media reputations of public agencies are related to their likelihood of experiencing structural reforms. This paper contributes to the literatures on organizational reform/terminations and bureaucratic reputation in several ways.

First, the literatures on organizational reform and termination have shown how reforms are seen by political-administrative leaders as instruments to achieve certain pre-defined goals (structural-instrumental perspective) and/or as unlikely interruptions in between periods of organizational stability and incremental change (institutional perspective) (Christensen & Laegreid, 2007; Kuipers et al. 2018). At the level of antecedents, the focus has been on a series of structural, cultural or environmental variables that either insulate agencies from interventions or pressure/incentivize political-administrative actors to engage in reforms (MacCarthaigh & Roness, 2012). The role of organizational reputation – or: how public organizations are perceived by their stakeholders (Carpenter, 2001) – has remained unexplored. In order to make sense of public sector performance, political-administrative actors rely on information shortcuts in their decision-making processes (Nielsen & Moynihan, 2017). Most public sector activities are limited in the extent to which they can be captured by measurable performance outputs (Van Dooren et al., 2015). Reputation serves as a crucial information shortcut for actors to process information in the evaluation of the performance of public organizations (Bovens & 't Hart, 2016). Poor reputations and crises reflect, channel and amplify latent feelings of discontent with regard to the performance of an organization. Situations in which the organization's legitimacy is at stake set the stage for strategic reorientations and critical decisions about the future of the organization (Alink et al., 2001).

Second, the literature on bureaucratic reputation started with Daniel Carpenter's observation that organizations can forge autonomy by developing favorable reputations that are recognized in stakeholder networks (Carpenter, 2001). While case-based studies demonstrate how well-reputed agencies avoid political interventions, and while large N studies show a positive relation between agencies' general reputations and their levels of discretion, no studies to date have provided similar evidence for the relation between reputations and the likelihood to experience reforms.

Third, this study contributes methodologically by relating a validated approach to measuring reputation as reputational history (Salomonsen et al., 2021) – based on supervised machine learning techniques to detect sentiment in texts towards 15 agencies over a period of 15 years – to a dataset on the structural reforms experienced by these agencies (Belgian State Administration Database, e.g. Kleizen et al., 2018).

In the remainder of this paper, we first introduce the concept of structural reforms, and discuss theoretical insights on the antecedents of reforms. We then introduce a reputational perspective on reforms. After introducing the data and empirical approach, we present and discuss the results.

Theoretical framework

Defining and understanding structural reforms

Defining reforms

Reforms should be distinguished from changes. Reforms refer to "active and deliberate attempts by political and administrative leaders to change structural or cultural features of organizations" (T. Christensen et al., 2007, p. 122). Changes refer to "what happens in practice and may or may not be the result of reforms" (MacCarthaigh & Roness, 2012, p. 7). In this paper, we focus on a variety of structural reforms that can be measured consistently and reliably over time (see 'Data' section).

This conceptualization of structural reforms into measurable reform events follows from the literature on agency termination. The comprehensive study of structural reforms in the public sector has a long history, and started with Kaufman's (1976) study on the mortality of government organizations. From this work emerged a literature on the termination of government organizations (Boin et al., 2010; O. James et al., 2015; Kuipers et al., 2018; MacCarthaigh & Roness, 2012; Park, 2013). The termination literature has been struggling with the issue how to conceptually and empirically make sense of a variety of reforms which are initiated by political and administrative elites (Peters & Hogwood, 1991). These include splittings, secessions, mergers, and absorptions, but also movements of organizations vertically and horizontally within the state apparatus, as well as into or out of it (Kuipers et al., 2018). The name "termination" suggests a very specific (and far-reaching) understanding of structural reforms as complete endings. This strict conceptualization is indeed followed by some scholars who consider organizations to be terminated when they are actually disbanded and cease to exist in any form (O'Leary, 2015). However, other scholars take a more encompassing view on termination and treat most structural reforms as termination types or events (Box-Steffensmeier & Jones, 2004; O. James et al., 2015). In between these extremes, large variance exists as to how scholars consider (combinations of) concrete reform events (e.g. name change, structural or leadership change) as either signs of discontinuity (termination) or survival (continuity); decisions that tend to be made on pragmatic choice rather than theoretical underpinning with severe repercussions for the comparability of termination rates across studies (Kuipers et al., 2018; Park, 2013).

In an attempt to bring much-needed standardization and develop a classification of reform events that can travel across administrative systems, the literature on the mapping of the formal structure of administrations distinguishes between three types of reform events: birth (or founding) events, survival (or maintenance) events, and death (or termination) events, which are then each further subdivided into subtypes (Kleizen et al., 2018; MacCarthaigh et al., 2012; Rolland & Roness, 2012). The present study uses the terminology in this latter group of studies, and will use the neutral term 'reform' instead of terminations.

Understanding reforms

At an ontological level, scholars have explicitly or implicitly ascribed to different perspectives on how to understand organizational reforms. Building on the longstanding collaboration between Johan P. Olsen and James G. March in which they combined organization theory, democratic theory and the study of decision-making behavior in formal organizations, Christensen et al. (2007) distinguish between a structural-instrumental perspective and an institutional perspective. These perspectives differ in important respects.

A *structural-instrumental perspective* views organizations as malleable tools in the hands of political-administrative leaders, who formulate goals exogenously after which organizations develop suitable means to reach said goals. It assumes self-interested rational actors with fixed preferences and identities that calculate expected returns from alternative options as a basis for their actions (i.e. logic of consequence): "The organization's leaders or superior authorities will calculate the costs and benefits of the organization's existing form and other relevant forms and will recommend reforms in cases where alternatives are deemed to be better for achieving the organization's goals" (T. Christensen et al., 2007, pp. 133–134). Reviewing the literature on agency termination, Kuipers, Yesilkagit and Carroll (2018) identify a political control tradition (which is contrasted with an institutional legacy tradition, see further). The political control tradition sees organizational reforms foremost as political acts, which are consciously imposed to increase political control. According to Carpenter and Lewis (2004), there are two main rationales for political actors to increase control through structural reforms: either for reasons of effectiveness (as a result of large observable failures, or because new and better methods have arisen for achieving the agency's ends), or for reasons of moral hazard (agencies are seen as ideologically unfaithful).

An *institutional perspective* views organizations as institutions (with rules, values, norms and routines) in their own right, which exert an independent influence on actions and decision-making. It assumes that organizational actors act in accordance with their experiences of what has worked well in the past, or upon what feels fair, reasonable and acceptable (i.e. logic of appropriateness) (T. Christensen et al.,

2007, p. 3). The result is more organizational robustness, with long periods of stability that are interrupted by radical breaks, as well as more variety in whether and how reforms are implemented, and a strong impact of national and organizational historical-institutional context (T. Christensen et al., 2007). In the termination literature, the institutional legacy tradition sees reforms (or better: the lack of them) as expressions of the fact that public organizations have their own interests: if not pure continuation and stability, then at least an adaptation of reform proposals to existing routines, norms and values (Kuipers et al., 2018).

These perspectives offer alternative views on how organizational behavior can be understood, based on distinct ontological assumptions of human action that either favor (top-level) agency (structural-instrumental perspective) or the constraining role of institutional context (institutional perspective). This study does not intend to settle longstanding debates on the merits of these perspectives. According to Hay (2002), attempts to "resolve" these debates by appealing to evidence are problematic given that such exercises conflate the ontological and the empirical (any given set of empirical observations can be accounted for by structural-instrumental and institutional perspectives). These perspectives, therefore, should not be seen as competing approaches, but as "different languages by which ontological differences between contending accounts might be registered" (Hay, 2002, p. 91). For instance, the implementation of New Public Management reforms can be seen as an optimal solution to efficiency problems (structural-instrumental perspective) and/or as an act of conformance to international prevailing doctrines (T. Christensen et al., 2007). This is also the approach taken in this study: rather than attempting to directly measure and compare mechanisms that would correspond to either perspective, these perspectives – combined with insights from the reputation literature (see further in this paper) – are used to formulate arguments for the expected relation between negative reputations and reform likelihood. Before doing so, however, we present an overview of the main antecedents of structural reforms as observed in the literature.

Antecedents of organizational reforms

Antecedents of structural reforms have been studied at the structural, cultural and environmental level (Kuipers et al., 2018; MacCarthaigh et al., 2012; Rolland & Roness, 2012). Studies that include *structural explanations* of organizational reforms take the position that the design of public organizations significantly affects their survival chances (Lewis, 2002). "Public organizations that are 'endowed' with certain structural features – such as political insulation, a specific organizational structure, and statutory recognition – enjoy higher survival chances than those without these birth characteristics'' (Boin et al., 2010, p. 385). Important structural variables found to affect the durability of agencies are the arm's length' nature (or: formal-legal status) of organizations, including whether they are governed

by a board or not, their form of establishment (by formal law or not, sunset clause or not) and their task (see Kuipers et al., 2018 for an overview). It should be noted, however, that these expectations are mainly predicated on the United States' system of separation of powers. In parliamentary systems such as Belgium, government have much greater opportunities for reorganization of termination of agencies, and are therefore less bound by the commitments of previous governments (Bertelli & Sinclair, 2016; Dommett & Skelcher, 2014).

Cultural explanations are found at the organizational level in variables that reflect agencies' resilience and durability, or at the sectoral or national level in variables related to (a) the isomorphic forces of (inter)national or sectoral trends and reform doctrines that have a legitimizing force, or (b) national or sectoral norms, values and routines. First, the age of agencies is among the most prominent and oldest factors to be included in studies on agency survival (Kuipers et al., 2018). The longstanding and conventional wisdom states that "the older a bureau is, the less likely it is to die" (Downs, 1967, p. 20), given that age reflects the time an agency has had to develop ties with a clientele from whom it draws power and resources (Kaufman, 1976; Lowi, 1979; Simon et al., 1950). In the first years after creation, agencies may still be protected by the legislative support base that has propagated its existence and is still in office. After about a decade, however, this protective shield may have disappeared, right at the moment when other legislators have had time to learn about and evaluate its performance (Carpenter & Lewis, 2004). Second, scholars point at the occurrence of fads, fashions and myths. For instance, Bertels and Schulze-Gabrechten (2021) observe how denominations for organizational units increasingly travel across ministerial borders, which may indicate the influence of fashions and myths on the façade of organizations. Third, explanations are sought in the institutional context. Verhoest et al. (2010) and MacCarthaigh (2012) identify the policy field in which an agency is active as a critical factor that shapes its autonomy and interference from political leaders or parent departments.

Lastly, *environmental variables* relate to changes in the political, economic and social context of agencies. Studies have included such variables as political turnover, political climate (ideology), political support, parliamentary composition, fiscal stress, unemployment rate, social demand for reform (Carpenter & Lewis, 2004; Kuipers et al., 2018; Lewis, 2002; Park, 2013). For instance, Park (2013) observes that the survival of agencies largely depends on external factors: termination is more likely during political turnover (see also Lewis, 2002), when presidential power is maximized and when social demands for reform are mature. Agencies are frequently targeted for political or ideological reasons, as incoming legislators and governments seek to bring the structure of an administration in line with their preferences (Carpenter & Lewis, 2004). This argument especially holds when the incoming (majority) government is either more right-wing (and more in favor of slimming down the public sector) or of a different political color than the organization's political creator and/or

administrative leadership (Kuipers et al., 2018). Lastly, the termination of agencies is particularly likely in times of fiscal stress, when competition among agencies for scarce resources is greatest (Kaufman, 1976), and when public attention for agencies increases (Bertelli & Sinclair, 2015; Park, 2013).

A reputational perspective on structural reforms

In this segment, we rely on the aforementioned ontological and explanatory frameworks to theorize and test the role of a hitherto neglected factor: organizational reputation.

Organizational reputation is defined as a set of beliefs about an organization's capacities, intentions, history, and mission that are embedded in a network of multiple audiences (Carpenter, 2010, p. 33). An organization's reputation is reflected in the perceptions of its internal and external audiences, which are diverse and may change over time (Maor, 2016). Reputations are built as organizations demonstrate that they have a unique capacity to create solutions and provide services (Carpenter, 2001). Strong reputations are crucial intangible assets for contemporary organizations, used "to generate public support, to achieve delegated autonomy and discretion from politicians, to protect the agency from political attack, and to recruit and retain valued employees" (Carpenter, 2002, p. 491). Scholars particularly point at the importance of reputation in the public sector, as public dissatisfaction and distrust with government is on the rise (Waeraas & Byrkjeflot, 2012). In the public sector, objective performance information that may counter negative stereotypes is difficult to collect (Carpenter & Lewis, 2004; Van Dooren et al., 2015), and often insufficient to remove existing biases (Hvidman & Andersen, 2016; Marvel, 2015), which further testifies to the importance of reputations rooted in subjective perceptions.

In the reputation literature, the assumption that favorable reputations protect agencies from political interventions is widespread. Bureaucratic actors are considered to spend much of their time on building and protecting a reputation that allows them to gain and maintain autonomy (Carpenter, 2001; Rourke, 1984; Wilson, 1989), and that serves as a protective shield against hostile audiences (Hood, 2013). Their reputation may greatly affect agencies' relationships with their elective and judicial overseers as a basis for autonomy (Carpenter & Krause, 2012). In the words of Carpenter (2001, p. 14): "Only when politicians and broad portions of the twentieth-century American public became convinced that some bureaucracies could provide unique and efficient public services, create new and valuable programs, and claim the allegiance of diverse coalitions of previously skeptical citizens did bureaucratic autonomy emerge". Carpenter (2001) proceeds with historical case studies of the Department of Agriculture and the Post Office Department in the United States to show how these departments' reputations gave these organizations significant autonomy and credibility towards legislators. In a rare effort to substantiate these mechanisms in a large N design, MacDonald and

Franko (2007) and MacDonald (2010) demonstrate that agencies with high grades in the Federal Performance Project are less likely to experience legislators' willingness to adopt limitation riders as a mechanism to rein in administrative discretion. While these studies show a positive relation between and agency's general reputation and its levels of discretion, no studies to date have provided similar evidence for the relation between reputations and the likelihood to experience reforms.

A reputational perspective informs the abovementioned structural-instrumental and institutional perspectives on organizational reforms. The structural-instrumental perspective calls attention to the dominant role of political-administrative elites in deciding on organizational reforms as an instrument to improve organizational outcomes and/or political control: "The question of reform and change will, in the first instance, be influenced by these people's goals and perceptions of the situation at hand" (T. Christensen & Laegreid, 2007, p. 133). To understand the likelihood of reform, therefore, is to understand the perceptions of political-administrative decision-makers.

The reputation of agencies is assumed to play an important role, because it provides decision-makers with information – albeit irregular and potentially biased (Carpenter & Lewis, 2004; Wilson, 1989) – about organizational performance. Most public sector activities are limited in the extent to which they can be captured by measurable performance outputs (Van Dooren et al., 2015). In order to make sense of public sector performance, political-administrative actors rely on information shortcuts in their decision-making processes (Nielsen & Moynihan, 2017). Reputation serves as a crucial information shortcut for actors to process information in the evaluation of the performance of public organizations (Bovens & 't Hart, 2016). In addition, reputations reflect the extent to which an organization is considered legitimate by its environment (Alink et al., 2001). A lack of legitimacy, regardless of the organization's actual performance, may be sufficient to trigger political sanctions because the mere *appearance* of administrative failure imposes costs on politicians (Wilson, 1989). The harsh repercussions of negative framing of government activities, coupled to the bias of the news media towards negative news (Boon et al., 2018), has even led scholars to recommend to public organizations a general hesitancy when it comes to searching the spotlights or propagating reputational excellence (Luoma-aho, 2007).

The responsiveness of political (and administrative) elites to media attention has been empirically demonstrated in different literatures. First, the notion that political elites use the media as an information source for their actions has been consistently confirmed in the literature on political agenda-setting, which has over a period of 50 years produced a strong body of empirical evidence on the impact of the media on the issues that political actors dedicate attention to (i.e. first-level agenda-setting) as well as how political actors think about these issues (i.e. second-level agenda-setting)

(McCombs & Shaw, 1972; Vliegenthart et al., 2016). McCombs et al. (2014) argue that the role of the media is enhanced when individuals deal with issues that are both of high personal relevance and when desired information is lacking. These conditions fit well with the general control problem of political principals vis-à-vis agencies that perform tasks of high public importance with substantial public funding, with high levels of autonomy and distance from direct political control.

Second, the public accountability literature is increasingly recognizing the role of the media throughout the accountability process of agencies. The media has a dual role, both as an accountability forum in their own right when the media independently calls attention to problems (e.g. through investigative journalism) and as a conduit for the opinions of other stakeholders (Jacobs & Schillemans, 2015). In addition, greater media attention indicates public salience and political demand (Carpenter, 2002). Both political actors and bureaucrats mention the media as a highly important external contingency in their daily work (Cook, 2006; Schillemans & Pierre, 2016), not least because of its potential disrupting impact on political-administrative relations and power (Carpenter & Lewis, 2004). Koop (2011) finds that agencies which deal with more salient issues are made more politically accountable. Bertelli and Sinclair (2015) show that agencies are less likely to face far-reaching reforms when they were salient in newspapers popular with the government's core supporters. Their findings suggest that political actors signal policy commitment to groups of voters, by maintaining or withdrawing independence from agencies depending on their media salience in particular newspapers. We therefore suggest the following hypothesis:

> *Hypothesis: Agencies with a more negative reputational history will have a higher likelihood to experience structural reforms.*

A reputational perspective also informs the institutional perspective on reforms, which tends to stress how reforms need to be accepted by, and translated to, organizational environments in order to be successful (Kuipers et al., 2018). Stability and incremental change is considered to be the dominant state for organizations, though this state of relative equilibrium may be punctuated by abrupt and powerful upheavals. Reforms are most likely to occur at so-called "critical junctures", which makes it important to clarify what precipitates these junctures (T. Christensen et al., 2007). The question becomes what causes these junctures develop. The reputation – or: accumulation of societal signals about the perceived performance of agencies – is again assumed to play an important role. Institutional scholars relate the likelihood of reforms to a *perceived* misfit between, on the one hand, an organization's identity, norms and routines and, on the other hand, the situation and environment it faces (T. Christensen et al., 2007). The role of perceptions and the perceived appropriateness of the current organization places reputation at center stage. Accumulated negative reputational signals may

serve as potent forces to overcome organizational inertia. In the private sector, numerous studies show how organizations tend to alter their activities in response to performance declines (see Romanelli & Tushman, 1994 for an overview). Indications of negative performance give change proponents the necessary arguments and power to overcome widespread resistance to change, and pursue (radical) reforms. Similar mechanisms seem to occur in the public sector. Poor reputations and crises reflect, channel and amplify latent feelings of discontent with regard to the performance of an organization. Situations in which the organization's legitimacy is at stake set the stage for strategic reorientations and critical decisions about the future of the organization (Alink et al., 2001).

Lastly, reputation scholars transcend the dominant focus of, on the one hand, the structural-instrumental perspective on reforms as instruments of political control and, on the other hand, the institutional perspective on bureaucratic actors as largely change-averse translators of reform propositions. The reputation literature strongly recognizes the role of bureaucratic actors as strategic actors in politics, who's relations with political actors can be characterized as dynamic, transactional and reputation-driven (Busuioc & Lodge, 2016; Carpenter & Krause, 2015). The structural-instrumental perspective on reforms – as a tool to reach optimal outcomes – may extend to bureaucratic actors whose self-interest and calculative behavior may drive reforms (J. G. Christensen, 1997; Dommett & Skelcher, 2014). Not unlike political elites, administrative elites are expected to be more likely to engage in reforms in response to negative reputational signals: if not to implement substantive changes to criticized aspects of their functioning, then at least as a symbolic act to demonstrate responsiveness to societal demands or as a strategy to avoid political blame. For instance, Alink et al. (2001) mention how crises may force actors to contemplate the possibility of outside intervention in order to avoid political blame avoidance. Accountability scholars show that bureaucratic actors engage in voluntary account-giving as a means to manage their reputation to external audiences (Busuioc & Lodge, 2016).

Data

Our study focuses on Flemish public agencies. Flanders is an autonomous region in the federalized system of Belgium. The Flemish government has its own parliament, cabinet, and public administration (consisting of departments and agencies). The Flemish government (and other regional governments) have equal legislative and executive powers, as decrees issues by the regional governments have the same legal standing as federal laws. Flanders should therefore be considered a full-fledged state for the competences under its remit (Verhoest et al., 2012).

Scholars have taken different approaches to measuring reputation (Bustos, 2021). The present study measures sentiment towards organizations in the traditional (written) news media (Gilad et al., 2015; Peci, 2021; Salomonsen et al., 2021). Though reputation has also been measured through more direct measures of audience perceptions (e.g. Capelos et al., 2016; Overman et al., 2020), the relevance of the media derives from its roles as (a) main source of mediated information about agencies , (b) active player in politics that informs and frames whether and how other stakeholders perceive agencies, and (c) institutional intermediary that gives voice and connects other audiences (Deephouse, 2016; Salomonsen et al., 2021; Strömbäck & Esser, 2014).

The media is not a neutral actor in its depiction of events and stories (Peci, 2021). Each event has a lot of potential attributes that may be emphasized to different extents (McCombs et al., 2014). Some of these are cognitive attributes, related to information about the characteristics of the object, and tell us what information to process. In a reputation context, these cognitive attributes may relate to the performative, technical, procedural and/or moral dimension of reputation (Carpenter, 2010; Rimkutė, 2020). Other attributes are affective, related to the tone of coverage (positive, negative or neutral sentiment), and tell us how to process information (Wu & Coleman, 2009). In this study, we focus on the sentiment of media coverage (positive, neutral, negative), which is highly relevant from a reputation viewpoint as it directly reflects the affective component of audience perceptions (Capelos et al., 2016). Stories with an explicit affective component may be particularly likely to affect reader perceptions and behaviors, given that affect triggers particular cognitive processes (related to the interpretation of issues, and attribution of responsibility) (Hood, 2013; Scheufele, 2000).

This study relies on machine learning methods to gather and code sentiment to public agencies in news media articles. Machine learning methods can broadly be categorized in two groups: *unsupervised techniques* which are primarily used to uncover latent patterns in data when there is no measured outcome; and *supervised techniques* which seek to replicate human coding on a known output. Machine learning methods can further be categorized into two types of prediction problems (James et al., 2017): *regression problems* involve the prediction of continuous variables, and *classification problems* rely on predicted probabilities to predict categorical variables. The present paper uses SML to code sentiment towards public organizations in text (Anastasopoulos & Whitford, 2018). As sentiment is coded as a set of classes (neutral, positive, negative), our research can be framed as a multiclass classification problem.

Main independent variable: Reputational history

Phase 1 Human coding

According to Anastasopoulos and Whitford (2018), models on data produced by expert coders perform significantly better than data produced by nonexperts. Most of the manual coding in the present study results from a related project in which sentiment towards the studied organizations in newspapers was hand-coded by a team of experts (see Salomonsen et al., 2021).

A text was coded as:

- 'neutral' if the agency, its action, or non-action are merely mentioned and/or described in neutral terms. No positive or negative wording or framing is used in the description;
- 'negative' if the agency was assigned responsibility for a negative incident, either through a description of the causal link between the agency's (in)action and the incident or when the agency's (in)actions – following an incident, or more generally – are framed negatively;
- 'positive' if the agency was assigned responsibility for a positive incident, either through a description of the causal link between the agency's (in)action and the incident or when the agency's (in)actions – following an incident, or more generally – are framed positively.

Our paper involves a relatively large variation in organizational settings. As a result, the hand-coded sample needed to be sufficiently large so that it reflected the large variety in words that may be used to reflect sentiment in these different settings. To further increase the size of the sample, additional texts were coded by one of the authors who is an expert in the field and was also part of the initial data coding. In total, 4916 texts were coded (1533 with negative sentiment; 1327 with positive sentiment; 2056 with neutral sentiment).

Phase 2 train and test

The texts used to predict sentiment were first preprocessed to make them suitable for machine learning. Using the Pattern package and sklearn package in Python (De Smedt & Daelemans, 2012; Pedregosa et al., 2011), a pipeline was built to tokenize and lowercase each document, and remove stop-words. The cleaned texts were then transformed into a *document-feature matrix*. Here, the rows contain the documents and the columns contain vectors that correspond to the features (i.e. word unigrams and bigrams) in all documents. Every cell contains the count of how often a feature occurs in every document. The end result is a numeric representation of the text of each document, which can be used for analysis.

Turning next to the actual training and testing, it is important to point out that different algorithms exist for classifying text, each with their own set of parameters that should be optimized during training. In order to find the most optimal algorithm and parameter setting, a procedure called *grid search cross validation* was used. "Grid search" refers to the search process of the procedure, in which the entire grid of possible combinations of parameters is searched and used for model training. "Cross validation" refers to the repeated evaluation of each model's performance on unseen data.

A typical feature of SLM is that data is split into train data (used for model building) and test data (used for model evaluation). In order to avoid overfitting, models should never be evaluated on data they have already seen during training. In this paper, we apply 10-fold cross-validation.[1] Data is split into 10 parts (or: folds) of equal size, after which a sequence of models is trained: the first model is built on folds 1-9 as training data and tested on the 10th fold; the second model uses the 2nd fold as test data and the remaining folds as training data. This process is repeated using folds 3, to 9 as test sets. In each run, once a model learns how to distinguish between texts using the training data, it is applied to the test data to make predictions about texts that the model has not seen. The performance of the machine-generated predictions are assessed by comparing the predictions against the human-coded data as a baseline (Anastasopoulos & Whitford, 2018). Each of the splits yield performance metrics, which averaged in an overall score (Müller & Guido, 2018).

Different performance metrics can be used to evaluate models. *precision* measures how many of the samples predicted in each class (negative, neutral, positive) correspond to the true labels; *recall* measures how many of the true samples in each class are captured by the predictions; the *f1-score*, lastly, provides a harmonic mean of precision and recall. These metrics are derived from the confusion matrix (see Table 1). In the confusion matrix, each cell allows to compare the instances of predicted labels for a particular class (e.g. negative opinion) with the actual labels. From the confusion matrix, the precision, recall and f1-score of each model can be calculated.

---

[1] Müller and Guido (2018, p. 254) mention a critical benefit of using cross-validation instead of a single split into a training and a test set. For instance, with a single random split, you can get "lucky" when all examples that are hard to classify end up in the training set. The test set will only contain "easy" examples, resulting in performance that is unrealistically high. When using cross-validation, each example will be in the test set exactly once: each example is in one of the folds, and each fold is the test set once. Therefore, the model needs to generalize well to all of the samples in the dataset for all of the cross-validation scores (and their mean) to be high.

Table 1: Confusion matrix

True classes	Predicted classes		
	Negative	Neutral	Positive
Negative	965	420	148
Neutral	300	1511	245
Positive	173	335	819

Table 2 performance statistics

	Precision	Recall	F1-score	Support
Negative	0.67	0.63	0.65	1533
Neutral	0.70	0.73	0.70	2056
Positive	0.68	0.62	0.65	1327
Accuracy			0.67	4916
Macro average	0.67	0.66	0.66	4916
Weighted average	0.67	0.67	0.67	4916

The end product of a Grid Search cross validation procedure is a "best model" (i.e. with optimized generalization performance estimated using only training data). The best performing model was the support vector model with a weighted average precision, recall, and f1-score of 67% (see Table 2).

Table 3 gives some of the features (words) that are associated (positive features) with the different output classes (negative opinion, neutral opinion, positive opinion). As this is not a dictionary-based study, these features are only indicative of the more complex patterns of words that the algorithm actually uses in its predictions.

Table 3: Feature analysis (translated from Dutch)

Negative opinion	[critical; insufficient; not good; dissatisfaction; complaints; accidents; error; responsibility; politics]
Neutral opinion	[circumstances; possibility; end; participants; come; property; need; to place; administrative]
Positive opinion	[success; good; safe; improve; thanks to; satisfied; support from; happy; strong; fruits;; pleased; interest; nice; proud; ideal; perfect; positive; comfort]

Phase 3 Predictions

After the optimal algorithm was selected, a new model was trained on the entire annotated data. For each observation, the selected text was passed through the classifier and labeled "neutral", "positive" or "negative". Table 4 shows the percentages of neutral, positive and negative texts for each organization.

Table 4: Descriptive overview of organizations and their media attention

Organization	Number of texts	% neutral	% positive	% negative
Flemish Public Employment Service	19503	78%	6%	16%
Sports Flanders	11298	83%	6%	10%
Child and Family	9828	78%	4%	19%
Public Waste Agency of Flanders	5729	91%	2%	7%
Flemish Environment Agency	3513	86%	3%	11%
Flemish Land Agency	2068	89%	3%	8%
Agency for Care and Health	1721	93%	1%	6%
Flemish Regulator for Energy and Gas	870	76%	4%	20%
Research Institute—Nature and Forest	686	83%	2%	15%
Flemish Agency for Social Housing	382	84%	3%	13%
Flemish Agency for Persons with Disabilities	367	36%	1%	62%
Agency for Entrepreneurship	335	70%	17%	13%
Institute for Science and Technology	149	64%	26%	10%
Agency for Infrastructure in Education	755	85%	7%	8%
Youth welfare agency	256	59%	8%	34%
Agency for Nature and Forests	5558	71%	5%	24%
Total / average	**63018 (total)**	**77% (avg)**	**6% (avg)**	**17% (avg)**

Aggregating sentiment data into reputational history measure

To calculate a measure of reputational history, we rely on the Deephouse media favorability index (Deephouse, 2000), which is a refinement of the Janis–Fadner index (Janis & Fadner, 1943). The index has already been used to measure reputation (Salomonsen et al., 2021), and measures the relative proportion of texts with positive sentiment towards the agency to texts with negative sentiment towards the agency while controlling for the overall volume of articles. We rely on a 365-day period, each article being given equal weight in line with previous utilizations of the index. The formula is:

$$x = \begin{cases} \frac{(f^2 - fu)}{r^2}, & \text{if} \quad f > u \\ 0 & \text{if} \; f = u \\ \frac{(fu - u^2)}{r^2}, & \text{if} \quad u > f \end{cases}$$

where f = number of articles with positive sentiment towards the agency, u = number of articles with negative sentiment towards the agency, and r = the total number of articles. The theoretical range of the measure is −1 to 1, where "1" indicates that all of the articles in the past 365 days contained positive sentiment to the agency, and "−1" that all of the articles in the past 365 days contained negative sentiment to the agency. A 0 value indicates balance, and thus an equal number of articles praising or threatening the reputation of the agency. The measure also implies that mere mentions of the agency will result in the reputational history going toward 0 (the denominator increases, thus resulting in a smaller relative share of positive/negative articles towards the agency). Although other media reputation indices have been used (Zhang, 2016), this measure was selected because the inclusion of neutral articles in the dominator smoothens the trend of reputational history. Put differently: singular positive or negative opinions do not weigh as heavily in terms of their ability to change an agency's overall reputation in a given space of time. This smoothening effect fits well with general understandings of media reputations as relatively stable and incrementally evolving (Etter et al., 2017; Salomonsen et al., 2021).

Main dependent variable: structural reform likelihood

In order to examine the likelihood of being reformed we make use of the Belgian State Administration Database (BSAD). The BSAD includes all changes in formal organizational structure since the founding of an organization until its termination. These changes are coined structural reforms and are defined as those reforms that change the organizational boundaries in terms of units included, change the tasks attributed to the organization and/or change the structural embeddedness of the organization in the wider public sector (i.e. its legal form and ministerial portfolio) (see e.g. MacCarthaigh & Roness, 2012; Lægreid et al., 2010). The database is constructed using an analysis of laws, decrees and decisions of the Flemish government available through legal repositories, supplemented by annual reports, parliamentary documents, (annual) reports produced by governmental organizations and organizational websites to confirm the effects of legal documentation.

The database uses a similar structure as the Norwegian State Administration Database (NSAD, see http://www.nsd.uib.no/polsys/en/civilservice/). More precisely, a predefined categorization that classifies structural reforms in three main categories is used: reforms related to the founding of an organization, reforms related to the survival or maintenance of an organization, and reforms related

to the termination of an organization. As such, the database is based on an institutional legacy point of view (Dommett & Skelcher, 2014) whereby maintenance events act as a go-between between organization life and death (Kuipers et al., 2017). For each main category of structural reforms there are several sub-categories, including merger, splitting, secession, and absorption, as well as movement of organizations vertically and horizontally within the state apparatus and into or out of it (Rolland & Roness, 2011 p.405-407). We are primarily interested in the effects of structural reforms imposed during the lifetimes of organizations. Thus, we focus on maintenance events while leaving birth and death events beyond consideration. More precisely, we include the following events:

- Maintenance by secession
- Maintenance by absorption
- Maintenance by absorption of tasks from another governmental level
- New superior organization at the same level
- New form of affiliation/ legal form
- New superior organization and new form of legal form
- Maintenance by the adoption of new tasks

The BSAD hence allowed us to vividly and accurately track the structural reforms the organizations in Table 4 experienced during their lifetime. Our dependent is operationalized as a dummy variable, when an organization experienced a reform in a given year this dummy was set to 1. If it encountered no structural reforms in a given year is set to 0.

Control variables

Furthermore, we include a set of control variables in order to account for the effects of variables other than reputation related factors. Our dataset allows us to control for the following factors:

1) **Type** is included in order to examine the effects of agency type. This categorical variable distinguishes between units with no legal independence but some managerial autonomy (reference category), units with legal independence vested in public law, and private law agencies.

2) **Task** is included as a categorical variable with the following values: regulation (reference category), exercising other kinds of authority, general public services towards other public organizations and general public services towards end users.

3) Age (**Age**) measured as '2015 (last year in dataset) - founding date' is included as a continuous variable.

Descriptive statistics

Table 5 Descriptive statistics (N=176)

Variable	Mean	Sd.
Experiencing a reform	0,131	0,338
Reputational history	-0,023	0,137
Age	21,830	9,035
Type	Frequency	
Units with no legal independence but some managerial autonomy	22,73%	
Units with legal independence vested in public law	40,91%	
Private law agencies	36,36%	
Task	Frequency	
Regulation	23,86%	
Exercising other kinds of authority	26,14%	
General public services towards other public organizations	21,59%	
General public services towards end users	28,41%	
Year	2009,591	3,797

Table 6 Correlation matrix

Variable		(1)	(2)	(3)	(4)	(5)	(6)
Experiencing a reform	(1)	1					
Reputational history	(2)	0,123	1				
Age	(3)	0,0653	0,064	1			
Type	(4)	0,1529	0,1883	0,5153	1		
Task	(5)	0,0216	-0,0337	0,1876	-0,2449	1	
Year	(6)	-0,2297	-0,2211	-0,2824	-0,2424	0,0202	1

Table 5 shows summary statistics for the main variables, while in Table 6 the correlation matrix is presented. Table 5 indicates that, on average, 13% of organizations experienced a reform in a given year. Note however that this is an average and that the examined time period ranged from 2000-2015. Additionally, we also test for multicollinearity using the variance inflation factor. The mean VIF equals 1.33. The highest VIFs exist for Task (1.72) and Age (1.67). These values indicate that no collinearity exists between the variables.

Method and results

For a first exploration of the association between reputational history and the likelihood of experiencing a reform, we make use of a standard logit model (because the dependent variable – experiencing a reform or not – is operationalized as a dummy). In order to accurately interpret the logit results, we calculated odds ratios. For a unit increase in x_k, the odds of a lower outcome

compared with a higher outcome are changed by the factor exp ($-\beta_k$), holding all other variables constant. Interpreting the results of such a Maximum Likelihood model is consequently different compared to a standard OLS regression (see Long & Freese, 2006, for a more thorough discussion). We cluster the standard errors by year. In order to establish a causal relationship between reputational history and the likelihood of experiencing a reform, we include 2 lags (based on year) of reputation. Including reputation of the same year as the reform does not make sense as the Deephouse index captures the whole year, while the reform will have happened during that year (we assume that sufficient time – that is: at least a year – needs to pass for an accumulation of negative reputational signals to trigger reforms). The results are presented in Table 7.

Table 7 Logit Regression results

Experiencing a reform	Odds	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Reputational history (t-1)	51.225	172.242	1.17	0.242	0.070	37289.779	
Reputational history (t-2)	**0.082**	**0.124**	**-1.65**	**0.098**	**0.004**	**1.589**	*
Age	0.977	0.047	-0.48	0.631	0.890	1.073	
Type: no legal independence		.	.	.	.	.	
Type: legal independence	3.841	4.117	1.25	0.209	0.470	31.396	
Type: private law agency	**6.063**	**5.814**	**1.88**	**0.060**	**0.926**	**39.715**	*
Regulatory task		.	.	.	.	.	
Other authoritative task	1.786	1.669	0.62	0.535	0.286	11.150	
Service delivery to publ org.	1.298	0.966	0.35	0.726	0.302	5.582	
Service delivery to end users	2.380	1.396	1.48	0.139	0.754	7.511	
Constant	0.040	0.034	-3.81	0.000	0.008	0.210	***
Mean dependent var		0.131	SD dependent var			0.338	
Pseudo r-squared		0.078	Number of obs			176.000	
Chi-square		70.222	Prob > chi2			0.000	
Akaike crit. (AIC)		143.776	Bayesian crit. (BIC)			172.311	

*** p<0.01, ** p<0.05, * p<0.1

These results indicate that only the reputational history measure of 2 years before the reform and being a private law agency seem to significantly affect the likelihood of experiencing a reform. Although barely significant, it appears that organizations with more negative reputational histories (note that these are odds ratios: everything below 1 is negative) are more likely to experience a reform after two years. The reputational history index is not significant the year before the reform. A possible reason for this finding could be that reforms take time and that the effect of reputation is actually a delayed effect. In general, it is also important to note that this regression is still very explorative. Future analyses will preferably account for the panel structure of the data, while the pooled models will be expanded with extra control variables (e.g. policy field, size,…).

Conclusion

The aim of this study was to shed light on the relation between reputational history and the likelihood that agencies experience structural reforms. Even though the relation between reputation and high-level political-administrative interventions is widely assumed in the reputation literature, and even though a reputational perspective has the potential to address largely unaddressed questions in the literature on structural reforms. A structural-instrumental perspective on reforms sees political-administrative decision-makers as key instigators of reform, yet it remains largely unclear on the basis of which information these actors actually decide to engage in reforms. An institutional perspective points at the general robustness of public organizations to change, which may however be interrupted by radical breaks. Again, however, studies have remained largely silent on the factors that set these breaks (or: critical junctures) in motion.

This paper built on these perspectives to develop a series of arguments for the expected positive relation between negative reputational histories and structural reform likelihood. The core argument is that reputational provide political-administrative actors – with limited insight in the workings of arm's length organizations – with information about the perceived performance of these agencies. Whether these perceptions reflect factual performance or not, it was generally expected that an accumulation of negative reputational signals (as measured in the reputational history indicator) would incentivize reforms.

Our results show the merit of this proposition, yet they also urge for further analyses on the robustness of this finding. On the one hand, we observe that – unlike traditionally included factors such as organizational age, formal-legal status or task – the effect of reputational history (lag 2) is significant. The time gap of two years between negative reputations and the implementation of reforms is not surprising. After all, it takes time for negative signals to accumulate to the extent that it alerts political-administrative leaders to call for reforms, negotiate and announce and, finally, implement them. On the other hand, we should be careful with this finding. The effect between reputational history (lag 2) and reform likelihood is barely significant at the 0.1 level (p = 0.098). Future, more refined analyses will need to confirm the validity of this finding: by using a panel data structure, by including more control factors (such as policy field, or more factors related to the environment: political turnover, fiscal pressures…), by testing more complex relation such as interaction effects, etc. In addition, it will be interesting to explore whether the date of the *announcement* of a reform (as opposed to the *implementation*) leads to clearer results. Furthermore, the reputational history measure in this study (see also Salomonsen et al., 2021) was adopted because it brings together the positive and negative attention towards organizations relative to their overall attention over a period of time in one single

measure. One drawback of the measurement, however, may be that it treats positive and negative opinions as equal. It could be that negative articles play a more substantive role in opinion formation than positive articles (cf. trust literature in which it is found that distrust has different affective and behavioral implications than trust). Therefore, future analyses could use different measures of reputation, giving more weight to negative articles or including positive and negative accumulated sentiment as separate items in the regressions. Related, the analyses of how particular distributions of reputations (e.g. more constantly negative/positive vs. sudden peeks in positive/negative attention) are related to structural reforms may lead to interesting insights. Lastly, it will be interesting to use more in-depth qualitative methods to explore why certain organizations fit our expectations particularly well (for instance, the Flemish Agency for Persons with Disabilities) whereas others correspond to completely different models (for instance, the Institute for Science and Technology has one of the best average reputations yet was also intensely reformed).

References

Alink, F., Boin, A., & T'Hart, P. (2001). Institutional crises and reforms in policy sectors: The case of asylum policy in Europe. *Journal of European Public Policy*, *8*(2), 286–306. https://doi.org/10.1080/13501760151146487

Anastasopoulos, L. J., & Whitford, A. B. (2018). Machine Learning for Public Administration Research, With Application to Organizational Reputation. *Journal of Public Administration Research and Theory*. https://doi.org/10.1093/jopart/muy060

Bertelli, A. M., & Sinclair, J. A. (2015). Mass Administrative Reorganization, Media Attention, and the Paradox of Information. *Public Administration Review*, *75*(6), 855–866. https://doi.org/10.1111/puar.12396

Bertelli, A. M., & Sinclair, J. A. (2016). Democratic Accountability and the Politics of Mass Administrative Reorganization. *British Journal of Political Science*, *48*(3), 691–711. https://doi.org/10.1017/S0007123416000077

Bertels, J., & Schulze-Gabrechten, L. (2021). Mapping the black box of intraministerial organization: An analytical approach to explore structural diversity below the portfolio level. *Governance*, *34*(1), 171–189. https://doi.org/10.1111/gove.12486

Boin, A., Kuipers, S., & Steenbergen, M. (2010). The Life and Death of Public Organizations: A Question of Institutional Design? *Governance*, *23*(3), 385–410. https://doi.org/10.1111/j.1468-0491.2010.01487.x

Boon, J., Salomonsen, H., Verhoest, K., & Pedersen, M. Ø. (2018). Media and Bureaucratic Reputation: Exploring Media Biases in the Coverage of Public Agencies. In *In T. Bach and K. Wegrich (eds) The Blind Spots of Public Bureaucracy and the Politics of Non-Coordination* (pp. 171–192). Palgrave MacMillan.

Bovens, M., & 't Hart, P. (2016). Revisiting the study of policy failures. *Journal of European Public Policy*, *23*(5), 653–666. https://doi.org/10.1080/13501763.2015.1127273

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790874

Bustos, E. O. (2021). Organizational Reputation in the Public Administration: A Systematic Literature Review. *Public Administration Review*. https://doi.org/10.1111/puar.13363

Busuioc, E. M., & Lodge, M. (2016). The Reputational Basis of Public Accountability. *Governance*, *29*(2), 247–263. https://doi.org/10.1111/gove.12161

Capelos, T., Provost, C., Parouti, M., Barnett, J., Chenoweth, J., Fife-Schaw, C., & Kelay, T. (2016). Ingredients of institutional reputations and citizen engagement with regulators. *Regulation & Governance*, *10*(4), 350–367. https://doi.org/10.1111/rego.12097

Carpenter, D. (2001). *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies 1862-1928*. Princeton University Press.

Carpenter, D. (2002). Groups, the Media, Agency Waiting Costs, and FDA Drug Approval. *American Journal of Political Science*, *46*(3), 490–505.

Carpenter, D. (2010). *Reputation and power: Organizational image and pharmaceutical regulation at the FDA*. Princeton Univ. Press.

Carpenter, D., & Krause, G. A. (2012). Reputation and public administration. *Public Administration Review*, *72*(1), 26–32.

Carpenter, D., & Krause, G. A. (2015). Transactional Authority and Bureaucratic Politics. *Journal of Public Administration Research and Theory*, *25*(1), 5–25. https://doi.org/10.1093/jopart/muu012

Carpenter, D., & Lewis, D. E. (2004). Political Learning from Rare Events: Poisson Inference, Fiscal Constraints, and the Lifetime of Bureaus. *Political Analysis*, *12*, 201–232.

Christensen, J. G. (1997). Interpreting Administrative Change: Bureaucratic Self-Interest and Institutional Inheritance in Government. *Governance*, *10*(2), 143–174. https://doi.org/10.1111/0952-1895.351997035

Christensen, T., Laegreid, P., Roness, P. G., & Røvik, K. A. (2007). *Organization Theory and the Public Sector: Instrument, Culture and Myth*. Routledge.

Cook, T. E. (2006). The News Media as a Political Institution: Looking Backward and Looking Forward. *Political Communication*, *23*(2), 159–171. https://doi.org/10.1080/10584600600629711

De Smedt, T., & Daelemans, W. (2012). Pattern for Python. *Journal of Machine Learning Research*, *13*(66), 2063–2067.

Deephouse, D. L. (2000). Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories. *Journal of Management*, *26*(6), 1091–1112. https://doi.org/10.1177/014920630002600602

Deephouse, D. L. (2016). Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories: *Journal of Management*. https://doi.org/10.1177/014920630002600602

Dommett, K., & Skelcher, C. (2014). Opening the Black Box of Administrative Reform: A Strategic-Relational Analysis of Agency Responses to Termination Threats. *International Public Management Journal*, *17*(4), 540–563. https://doi.org/10.1080/10967494.2014.958805

Downs, A. (1967). *Inside Bureaucracy*. Little Brown.

Etter, M., Ravasi, D., & Colleoni, E. (2017). Social Media and the Formation of Organizational Reputation. *Academy of Management Review*, *44*(1), 28–52. https://doi.org/10.5465/amr.2014.0280

Gilad, S., Maor, M., & Bloom, P. B.-N. (2015). Organizational Reputation, the Content of Public Allegations, and Regulatory Communication. *Journal of Public Administration Research and Theory*, *25*(2), 451–478. https://doi.org/10.1093/jopart/mut041

Hay, C. (2002). *Political Analysis: A Critical Introduction*. Palgrave Macmillan.

Hood, C. (2013). *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton University Press. http://press.princeton.edu/titles/9353.html

Hvidman, U., & Andersen, S. C. (2016). Perceptions of Public and Private Performance: Evidence from a Survey Experiment. *Public Administration Review*, *76*(1), 111–120. https://doi.org/10.1111/puar.12441

Jacobs, S., & Schillemans, T. (2015). Media and public accountability: Typology and exploration. *Policy & Politics*, *44*(1), 23–40.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R*. Springer.

James, O., Petrovsky, N., Moseley, A., & Boyne, G. A. (2015). The Politics of Agency Death: Ministers and the Survival of Government Agencies in a Parliamentary System. *British Journal of Political Science*, *FirstView*, 1–22. https://doi.org/10.1017/S0007123414000477

Janis, I. L., & Fadner, R. H. (1943). A coefficient of imbalance for content analysis. *Psychometrika*, *8*(2), 105–119. https://doi.org/10.1007/BF02288695

Kaufman, H. (1976). *Are government organizations immortal?* Brookings Institution.

Kleizen, B., Verhoest, K., & Wynen, J. (2018). Structural reform histories and perceptions of organizational autonomy: Do senior managers perceive less strategic policy autonomy when faced with frequent and intense restructuring? *Public Administration*, *96*(2), 349–367. https://doi.org/10.1111/padm.12399

Koop, C. (2011). Explaining the Accountability of Independent Agencies: The Importance of Political Salience. *Journal of Public Policy*, *31*(02), 209–234. https://doi.org/10.1017/S0143814X11000080

Kuipers, S., Yesilkagit, K., & Carroll, B. (2018). Coming to Terms with Termination of Public Organizations. *Public Organization Review*, *18*(2), 263–278. https://doi.org/10.1007/s11115-017-0376-4

Lewis, D. E. (2002). The Politics of Agency Termination: Confronting the Myth of Agency Immortality. *The Journal of Politics*, *64*(1), 89–107.

Lowi, T. J. (1979). *The End of Liberalism: The Second Republic of the United States*. Norton.

Luoma-aho, V. L. (2007). Neutral Reputation and Public Sector Organizations. *Corporate Reputation Review*, *10*(2). https://doi.org/10.1057/palgrave.crr.1550043

MacCarthaigh, M. (2012). Organisational Change in Irish Public Administration. *International Journal of Public Administration*, *35*(12), 795–807.

MacCarthaigh, M., & Roness, P. G. (2012). Analyzing Longitudinal Continuity and Change in Public Sector Organizations. *International Journal of Public Administration*, *35*(12), 773–782. https://doi.org/10.1080/01900692.2012.715567

MacCarthaigh, M., Roness, P. G., & Sarapuu, K. (2012). Mapping Public Sector Organizations: An Agenda for Future Research. *International Journal of Public Administration*, *35*(12), 844–851. https://doi.org/10.1080/01900692.2012.715568

MacDonald, J. A. (2010). Limitation Riders and Congressional Influence over Bureaucratic Policy Decisions. *American Political Science Review*, *104*(4), 766–782.

MacDonald, J. A., & Franko Jr., W. W. (2007). Bureaucratic Capacity and Bureaucratic Discretion Does Congress Tie Policy Authority to Performance? *American Politics Research*, *35*(6), 790–807.

Maor, M. (2016). Missing Areas in the Bureaucratic Reputation Framework. *Politics and Governance*, *4*(2), 80–90. https://doi.org/10.17645/pag.v4i2.570

Marvel, J. D. (2015). Public Opinion and Public Sector Performance: Are Individuals' Beliefs About Performance Evidence-Based or the Product of Anti–Public Sector Bias? *International Public Management Journal*, *18*(2), 209–227. https://doi.org/10.1080/10967494.2014.996627

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, *36*(2), 176–187. JSTOR.

McCombs, M. E., Shaw, D. L., & Weaver, D. H. (2014). New Directions in Agenda-Setting Theory and Research. *Mass Communication and Society*, *17*(6), 781–802. https://doi.org/10.1080/15205436.2014.964871

Müller, A. C., & Guido, S. (2018). *Introduction to Machine Learning With Python*. O'Reilly.

Nielsen, P. A., & Moynihan, D. P. (2017). How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy. *Journal of Public Administration Research and Theory*, *27*(2), 269–283. https://doi.org/10.1093/jopart/muw060

O'Leary, C. (2015). Agency Termination in the UK: What Explains the 'Bonfire of the Quangos'? *West European Politics*, *38*(6), 1327–1344. https://doi.org/10.1080/01402382.2015.1010273

Overman, S., Busuioc, M., & Wood, M. (2020). A Multidimensional Reputation Barometer for Public Agencies: A Validated Instrument. *Public Administration Review*, *n/a*(n/a). https://doi.org/10.1111/puar.13158

Park, S. (2013). What Causes the Death of Public Sector Organizations? Understanding Structural Changes and Continuities in Korean Quangos. *International Public Management Journal*, *16*(3), 413–437. https://doi.org/10.1080/10967494.2013.825491

Peci, A. (2021). Agencies in the news? Public agencies' media evaluations in a low-trust context. *Governance.* https://doi.org/10.1111/gove.12579

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peters, B. G., & Hogwood, B. W. (1991). Applying population ecology models to public organizations. In *In Research in Public Administration (Ed. J.L. Perry)*. JAI Press.

Rimkutė, D. (2020). Building organizational reputation in the European regulatory state: An analysis of EU agencies' communications. *Governance*, *33*(2), 385–406. https://doi.org/10.1111/gove.12438

Rolland, V. W., & Roness, P. G. (2012). Foundings and Terminations: Organizational Change in the Norwegian State Administration 1947–2011. *International Journal of Public Administration*, *35*(12), 783–794. https://doi.org/DOI: 10.1080/01900692.2012.715562

Romanelli, E., & Tushman, M. L. (1994). Organizational Transformation as Punctuated Equilibrium: An Empirical Test. *The Academy of Management Journal*, *37*(5), 1141–1166. https://doi.org/10.2307/256669

Rourke, F. E. (1984). *Bureaucracy, Politics, and Public Policy*. Little Brown.

Salomonsen, H. H., Boye, S., & Boon, J. (2021). Caught up or Protected by the Past? How Reputational Histories Matter for Agencies' Media Reputations. *Journal of Public Administration Research and Theory*, *muaa056*. https://doi.org/10.1093/jopart/muaa056

Scheufele, D. A. (2000). Agenda-Setting, Priming, and Framing Revisited: Another Look at Cognitive Effects of Political Communication. *Mass Communication and Society*, *3*(2–3), 297–316. https://doi.org/10.1207/S15327825MCS0323_07

Schillemans, T., & Pierre, J. (2016). Entangling and disentangling governance and the media. *Policy & Politics*, *44*(1), 1–8. https://doi.org/10.1332/147084415X14503518462782

Simon, H. A., Smithburg, D. W., & Thompson, V. A. (1950). *Public Administration*. Alfred A. Knopf.

Strömbäck, J., & Esser, F. (2014). Meditization of Politics: Towards a Theoretical Framework. In *Mediatization of Politics: Understanding the Transformation of Western Democracies* (pp. 3–28).

Van Dooren, W., Bouckaert, G., & Halligan, J. (2015). *Performance Management in the Public Sector (2nd edition)*. Routledge.

Verhoest, K., Roness, P. G., Verschuere, B., Rubecksen, K., & MacCarthaigh, M. (2010). *Autonomy and Control of State Agencies; Comparing States and Agencies*. Palgrave Macmillan.

Verhoest, K., Van Thiel, S., Bouckaert, G., & Laegreid, P. (2012). *Government agencies: Practices and lessons in 30 countries*. Palgrave Macmillan.

Vliegenthart, R., Walgrave, S., Baumgartner, F. R., Bevan, S., Breunig, C., Brouard, S., Bonafont, L. C., Grossman, E., Jennings, W., Mortensen, P. B., Palau, A. M., Sciarini, P., & Tresch, A. (2016). Do the media set the parliamentary agenda? A comparative study in seven countries. *European Journal of Political Research*, *55*(2), 283–301. https://doi.org/10.1111/1475-6765.12134

Waeraas, A., & Byrkjeflot, H. (2012). Public Sector Organizations and Reputation Management: Five Problems. *International Public Management Journal*, *15*(2), 186–206.

Wilson, J. Q. (1989). *Bureaucracy: What Government Agencies do and Why they do it*. Basic Books.

Wu, H. D., & Coleman, R. (2009). Advancing Agenda-Setting Theory: The Comparative Strength and New Contingent Conditions of the Two Levels of Agenda-Setting Effects. *Journalism & Mass Communication Quarterly*, *86*(4), 775–789. https://doi.org/10.1177/107769900908600404

Zhang, X. (2016). Measuring Media Reputation: A Test of the Construct Validity and Predictive Power of Seven Measures. *Journalism & Mass Communication Quarterly*, *93*(4), 884–905. https://doi.org/10.1177/1077699015610065