

A reputational perspective on structural reforms: how media reputations are related to the structural reform likelihood of public agencies

Peer-reviewed author version

BOON, Jan; Wynen , Jan; Verhoest, Koen; Daelemans, Walter & Lemmens , Jens (2024) A reputational perspective on structural reforms: how media reputations are related to the structural reform likelihood of public agencies. In: Journal of public administration research and theory,.

DOI: 10.1093/jopart/muae023

Handle: <http://hdl.handle.net/1942/44932>

# A reputational perspective on structural reforms:

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

Jan Boon\*,\*\*\*

[jan.boon@uhasselt.be](mailto:jan.boon@uhasselt.be)

Martelarenlaan 42 3500 Hasselt, Belgium

Jan Wynen\*\*,\*\*\*

[jan.wynen@uantwerpen.be](mailto:jan.wynen@uantwerpen.be)

Prinsstraat 13 2000 Antwerp, Belgium

Koen Verhoest\*\*\*

[koen.verhoest@uantwerpen.be](mailto:koen.verhoest@uantwerpen.be)

Sint - Jacobstraat 2 – 4 2000 Antwerp, Belgium

Walter Daelemans\*\*\*\*

[walter.daelemans@uantwerpen.be](mailto:walter.daelemans@uantwerpen.be)

Lange Winkelstraat 40 2000 Antwerp, Belgium

Jens Lemmens\*\*\*\*

[jens.lemmens@uantwerpen.be](mailto:jens.lemmens@uantwerpen.be)

Lange Winkelstraat 40 2000 Antwerp, Belgium

\*School of Social Sciences, University of Hasselt

\*\*Department of Management, University of Antwerp

\*\*\* Politics & Public Governance, Department of Political Sciences, University of Antwerp

\*\*\*\*Centre for Computational Linguistics and Psycholinguistics, University of Antwerp

### Abstract:

Despite recurrent observations that media reputations of agencies matter to understand their reform experiences, no studies have theorized and tested the role of sentiment. This study uses novel and advanced BERT language models to detect attributions of responsibility for positive/negative outcomes in media coverage towards 14 Flemish (Belgian) agencies between 2000-2015 through supervised machine learning, and connects these data to the Belgian State Administration Database on the structural reforms these agencies experienced. Our results reflect an inverted U-shaped relationship: more negative reputations increase the reform likelihood of agencies, yet up to a certain point at which the reform likelihood drops again. Variations in positive and neutral reputational signals do not impact the reform likelihood of agencies. Our study contributes to understanding the role of reputation as an antecedent of structural reforms. Complementing and enriching existing perspectives, the paper shows how the sentiment in reputational signals accumulates and informs political-administrative decision-makers to engage in structural reforms.

Key words: organizational reputation; structural reform; supervised machine learning; BERT language model

## Introduction

Through structural reforms, governments seek to design public agencies that will stand the test of their environment (Kleizen, Verhoest, and Wynen 2018). Yet reforms are also resource- and time-consuming, and tend to lead to disappointing results (Wynen, Verhoest, and Kleizen 2019). It is therefore important to understand why political-administrative actors engage in reforms. While the question elicited substantial research (Christensen et al., 2007; Kuipers et al., 2018; MacCarthaigh et al., 2012), studies tend to restrict their attention to structural, cultural and environmental factors that make public organizations more or less likely to experience reform (MacCarthaigh and Roness, 2012). More broadly, theoretical perspectives tend to see reforms as instruments for political-administrative leaders 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 and Laegreid, 2007; Kuipers et al. 2018).

In this study we shed light on the role of agencies' media reputations, asking *to what extent media salience and valence matter for agencies' likelihood to experience structural reforms*. Reputation scholars point at reputational concerns of diverse actors as key drivers of a wide variety political and organizational behaviors (Busuioc and Lodge 2016; Wæraas and Maor 2015). In order to make sense of public sector performance, political-administrative actors rely on information shortcuts in their decision-making processes (Nielsen and Moynihan 2017). Most public sector activities are limited in the extent to which they can be captured by measurable performance outputs (Van Dooren, Bouckaert, and Halligan 2015). Reputation serves as a crucial information shortcut for actors to process information in the evaluation of the performance of public organizations (Bovens and 't Hart 2016). Favorable reputations are intangible assets for agencies to build and secure autonomy from politics (Carpenter 2001; Wæraas and Maor 2015). Poor reputations 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, Boin, and T'Hart 2001).

This study seeks to further unravel the interplay between reputation theory and the reform literature, a nascent yet promising coupling (Chen, Christensen, and Ma 2023; Bustos 2021). Reputation theory has a proven track record with explaining political-administrative behaviors through reputational considerations (Wæraas and Maor 2015; Busuioc and Lodge 2016). While the reputation literature built on Carpenter's (2001;2010) case-based insights that conceived of reputation as an autonomy-enhancing asset that insulates agencies from political interference, the link between reputation and structural reforms – arguably among the most invasive interventions – has received surprisingly little

attention. Of 53 articles focused on the outcomes of organizational reputation, Bustos (2021) identified a surprisingly low number that considered the effect of reputation on autonomy (N = 3) or bureaucratic power (N = 2) – and none directly examined structural reforms.

In a recent study, Chen, Christensen and Ma (2023) were the first to point out the potential of reputation theory to better understand agency terminations. Building on previous literature showing how media attention in general matters for terminations (Bertelli and Sinclair 2015), they coupled these insights to theorize the role of the different (performative, moral, technical, procedural) reputational dimensions (Carpenter 2010). This study builds on these studies in two ways. First, by exploring whether the relevance of general media attention holds when using a broader measure of structural reforms than agency terminations, and in a different setting than hereto studied. Second, it theorizes and tests the role of positive, negative and neutral media reputations. While the relevance of including a sentiment measure of reputation has been criticized (Bertelli and Sinclair 2015; Chen, Christensen, and Ma 2023), previous studies have argued for its non-inclusion on mere theoretical grounds. With arguments for its theoretical relevance also existing, studies are needed that outline the theoretical (dis)advantages of sentiment-based reputation, after which its explanatory value is empirically (dis)confirmed.

Particularly in the political science tradition (Wæraas and Maor 2015), reputation initially tended to be studied as a multidimensional phenomenon that considers agencies' reputation management efforts, and audiences' evaluation efforts, alongside four cognitive dimensions (Carpenter 2010). Gradually and increasingly, scholars are recognizing the affective component of reputations. While some scholars have linked the moral dimension to the affective part of perception formation (Tom Christensen and Gornitzka 2019), others have called for a more encompassing treatment of affect as a particular component of reputation (Boon, Salomonsen, and Verhoest 2021; Capelos et al. 2016). In a factor analysis, Lee and Van Ryzin (2018) found negatively worded dimensions to all load on a single factor, which provides some empirical support for the idea of a distinct affective reputation component. Recent studies have documented how the sentiment of media coverage is a key determinant to understand legislators' questioning behavior about agencies (Boon et al. 2023), an insight that finds theoretical backing by political agenda-setting theory (Vliegenthart et al. 2016; McCombs, Shaw, and Weaver 2014) and negativity bias scholarship (Nielsen and Moynihan 2017; George et al. 2020). We argue that a sentiment perspective has the potential to contribute substantially to extant theory. It allows to explore the distinct impacts of negative and positive media reputations, with the former arguably more likely to elicit reforms than the latter. Likewise, it can shed light on the accumulating effects of certain signals; that is: the level of negativity agencies can afford before a certain reform

appetite is triggered, as well the level of positivity that is required to protect agencies from structural reforms.

Lastly, this study contributes methodologically to a growing body of research that applies machine learning techniques to topics of public management and administration (e.g. Anastasopoulos and Whitford 2018; Hollibaugh 2019; Chen, Christensen, and Ma 2023). Advanced supervised machine learning techniques – BERT language models – are used which are uniquely suited to take (organizational) context into account when making sentiment predictions. With BERT, we build an affect-based reputation measure for 14 organizations over a 15-year period (2000-2015) from the Flemish government (Belgium), detecting attributions of responsibility for positive/negative outcomes towards agencies in media coverage (Salomonsen, Boye, and Boon 2021). This measure is then coupled to a dataset on the structural reforms experienced by these agencies (Belgian State Administration Database, e.g. Kleizen et al., 2018).

## **Theoretical framework**

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, 33). 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, 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 and Byrkjeflot 2012). In the public sector, objective performance information that may counter negative stereotypes is difficult to collect (Carpenter and Lewis 2004; Van Dooren, Bouckaert, and Halligan 2015), and often insufficient to remove existing biases (Hvidman and Andersen 2016; Marvel 2015), which further testifies to the importance of reputations rooted in subjective perceptions.

The "embedded" (Carpenter 2010, 33) nature of reputation refers to a certain accumulation of reputational signals over a period of time, as beliefs about the agency are intricately woven into the collective understanding of the agency among its audiences. While there is some debate about the malleability of reputations, both in the sense of their endurance and their strategic potential (Boon 2022), there is a general agreement that particular reputational dynamics should be understood against the backdrop of a preceding reputational track record. For instance, Maor et al. (2013) find

that the responsiveness of the Israeli Banking Supervision Department in media articles depends on whether the article dealt with an issue on which the agency had developed a strong or weak reputation in the preceding years. Salomonsen et al. (2021) theorize and demonstrate the stickiness of reputations: agencies with more negative media reputations tend to attract more criticism in the following years, whereas the opposite is true for positively perceived agencies. This study, too, conceives of reputation as an asset (or constraint) that accumulates over time to impact structural reforms.

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 political and societal developments (Hood 2013). Their reputation may greatly affect agencies' relationships with their elective and judicial overseers as a basis for autonomy (Carpenter and Krause 2012).

Concerning structural reforms, a reputational perspective informs well-established perspectives that center on structural-instrumental and institutional factors (Kuipers, Yesilkagit, and Carroll 2018; Chen, Christensen, and Ma 2023), while also providing its own line of argument (cf. Table 1). This study does not intend to settle longstanding debates on the merits of these perspectives, and will include control variables of both these perspectives in the analyses. Our main theoretical interest is the role of organizational reputation, and how it complements – and is informed by – existing theoretical perspectives. In this sense, our approach follows that of Olsen (1992) that integrates different perspectives in developing more complex and psychologically realist models of bureaucratic politics (Carpenter 2010; Wæraas and Maor 2015). As such, a reputational perspective has the potential to contribute to the theorization of structural reforms, yet has received surprisingly little attention.

[Insert Table 1 here]

The structural-instrumental perspective calls attention to the dominant role of political elites in deciding on organizational reforms as an instrument to improve organizational outcomes and/or political control (Chen, Christensen, and Ma 2019): "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" (Christensen and Laegreid, 2007, p. 133). To understand the likelihood of reform, therefore, is to understand the perceptions of political-administrative decision-makers (Chen, Christensen, and Ma 2023). The reputation of agencies is assumed to play an important role, because it provides decision-makers with information – albeit irregular and potentially biased (Carpenter and 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, Bouckaert, and Halligan 2015). In order to make sense of public sector performance, political-administrative actors rely on information shortcuts in their decision-making processes (Nielsen and Moynihan 2017). Reputation serves as a crucial information shortcut for actors to process information in the evaluation of the performance of public organizations (Bovens and 't Hart 2016).

The responsiveness of political (and administrative) elites to media attention has been empirically demonstrated in different literatures. Political agenda-setting scholarship 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 (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. Furthermore, 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 and Schillemans 2015). In addition, greater media attention indicates public salience and political demand (Carpenter 2002).

The question then becomes what political elites *do* when encountering agency-centered media attention. After all, a big leap exists between thinking about agencies and acting upon them through reform. Furthermore, alternative explanations for the media-reform nexus than the previous focusing on political responsiveness have been posited. Institutional scholarship (Fig. 1 arrow 2) assumes that organizational actors act in accordance with what has worked well in the past, or upon what feels fair, reasonable and acceptable (i.e. logic of appropriateness) (Christensen et al., 2007, p. 3). Organizational forces that promote stability and robustness will dominate media influences, which in general tends to be highly neutral in their coverage of agencies to begin with (Boon, Salomonsen, and Verhoest 2019). Furthermore, mediatization scholarship has demonstrated how agencies have adapted institutionally to the media's logic and demands, allowing them to pro- and reactively deal with media attention and potential disturbances (Fredriksson, Schillemans, and Pallas 2015). Reputation scholars (Fig. 1 arrow 3), furthermore, have argued that media attention may actually reflect agencies' coalition-building efforts and stakeholder support base. As agencies gain public visibility, political actors may be constrained in their efforts to impose public sanctions (Chen, Christensen, and Ma 2023), particularly so when media attention occurred in newspapers consumed by core supporters of the incumbent government (Bertelli and Sinclair 2015).

We follow these latter explanations, expecting that *as media attention for agencies goes up, these agencies will be less likely to face structural reforms (H1)*.

Next, we examine the role of sentiment: to what extent does it matter whether coverage about agencies is positive or negative? Several authors have argued for the irrelevance of including sentiment in studies on reputation effects – albeit on theoretical rather than empirical grounds. The argument is that while media coverage may be positive, negative or (most often) neutral, what governments learn about public perceptions from such stories will be idiosyncratic and difficult to translate in directional hypotheses about tone/sentiment (Bertelli and Sinclair 2015). Building on the agency that reputation-sensitive agencies have in turning (any) media attention into a mobilizing force, Chen Christensen and Ma (2023, p. 221) state that “coverage per se more important than the kind of coverage” and that “even a report on an agency concerning a scandal or crisis is good”.

Other scholars and theories, however, disagree with the all-exposure-is-good-exposure argument. Political agenda setting theory distinguishes between first- and second-level agenda setting. Whereas the former captures the idea that media attention attracts political attention, the latter states that how the media discusses (or frames) issues also spills over to the attitudes *and behaviors* of media-consuming actors (McCombs, Shaw, and Weaver 2014). In addition, while sentiment in media about agencies can be perceived idiosyncratically by different audiences depending on psychological and institutional factors (Chen, Christensen, and Ma 2023; Bertelli and Sinclair 2015), alternative theories suggest that as negative (or positive) signals *accumulate* the political and organizational cost of ignoring them rises. Attribution theory is concerned with how people use information to arrive at a causal explanation for a behavior or event as a guide for future action (Barnes and Henly 2018). Two main factors that individuals consider in their psychological evaluation of an actor’s performance are (a) the control an actor had over a positive or negative outcome and (b) the stability by which good or bad performance occurs (Weiner 1985). When attributions accumulate to a positive overall impression (or: reputation), a “Halo-effect” emerges in which stakeholders view a negative incident as an anomaly; one misstep in a stable history of positive actions. When, however, negative coverage accumulates, the “Velcro-effect” refers to new incidents being ascribed to a stable pattern of poor performance that sticks to the organization (Salomonsen, Boye, and Boon 2021). Whereas “halos” protect organizations from unwanted interventions (such as resource-costly structural reforms), accumulated negativity is likely to trigger political-administrative reactions because the mere appearance of administrative failure imposes costs on politicians and agency leaders (Wilson 1989).

This latter mechanism also features in institutional theories on reform. While stability is considered to be the dominant state for organizations, 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). 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. Indications of negative performance give change proponents the necessary arguments and power to overcome widespread resistance to change, and pursue (radical) reforms. 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, Boin, and T’Hart 2001).

Such reorientation may also come from agencies themselves. The reputation literature (Fig. 1 arrow 3) strongly recognizes the role of bureaucratic actors as strategic actors in politics (Busuioc and Lodge 2016; Carpenter and 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 (Christensen, 1997; Dommett and 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. Reversely, positive reputations can be used as agency assets to shield an agency from political intervention through organizationally costly structural reforms (Carpenter, 2001).

Following this line of argument, we hypothesize that *as reputations are more positive, agencies are less likely to face structural reforms; whereas more negative reputations are more likely to lead to structural reforms*<sup>1</sup> (H2).

Lastly, our study allows to explore the relative strength of a negative vs. positive reputational track record in the media. So far, we discussed positive and negative reputations as two opposite sides of the same coin. Yet research convincingly shows that when it comes to influencing human – and particularly political – behavior, negative signals strongly outweigh positive signals. As people tend to prioritize preventing loss over obtaining gains, negative news will evoke a stronger attitudinal response

---

<sup>1</sup>See Data section: when we refer to “more negative” reputations, this means that the share of negative articles relative to all articles (positive, negative, neutral) is higher compared to other organizations and/or time periods. The same reasoning applies to “more positive” and “more neutral” reputations.

than positive news (Jonkman et al. 2020). Comparing the lingering effects of organizational reputations in the media, Salomonsen et al. (2021) observed that negative reputations tended to “stick” more than positive reputations. While positive reputations, too, had a lingering effect – this effect only took hold with levels of positivity that were difficult to attain or see in practice. 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 relative importance of negative news certainly seems to hold for politicians, who “must consider that they might be held responsible for their actions or inactions—or how these are played out in the media” (Strömbäck 2008, 239). Conflict, after all, is at the heart of politics (Schattschneider 1960). Politicians tend to focus more on controversial issues, which confront them with societal problems on which they might need to make their standpoint clear to the electorate. Thesen (2013) shows that the political agenda-setting effect of the news media is stronger when blame attributions are present. In an experimental set-up, Helfer and van Aelst (2020) find that negative news reports trigger politicians to take political action. Applying these insights to political attention for agencies through written questions, Boon et al. (2023) found that only negative media attention for agencies had an enduring impact on political attention for agencies in parliamentary questions. Bureaucrats, too, have been shown to be responsive to negativity. 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 and Lodge 2016).

Given these arguments, we hypothesize that *the reform-inducing effects of more negative reputations will be stronger than the reform-reducing effect of more positive reputations* (H3).

While we believe the hypothesis accurately represents the dominant assumed mechanisms in the literatures on reputation and political accountability, and hence serves as our guiding hypothesis, it is also worth pointing more succinctly to an alternative theoretical perspective for negative reputations. Some reputation scholars have argued that poor reputations, rather than societal signals to respond to through structural reforms, are rather stable facts of life for many public agencies. As Waeraas and Byrkjeflot (2012, 201) argue, “*public organizations that struggle with a poor reputation may not have their legitimacy questioned as long as their raison d’être is not challenged. Tax authorities serve as a good example. Few people like or admire them, or are willing to ascribe a strong reputation to them, yet most people would agree that they need to exist*”. Their line of reasoning may extend to the

likelihood of these agencies experiencing reforms as a function of their reputational track record. Given that negative reputations for many agencies may not challenge their legitimacy, the felt urge to be responsive to external criticisms through reforms may reduce as negativity becomes a normalized feature of their reputation. Despite the plausibility of this alternative perspective concerning the role of negative reputations, we will not formulate an explicit alternative hypothesis but rather reflect on the value of this reasoning vis-à-vis the guiding hypothesis in the results and discussion sections.

## **Data**

Our study focuses on 14 Flemish public agencies in the period 2000-2015. 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 issued 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). This focus on Flemish agencies, rather than the broader federalized system of Belgium, was chosen due to the distinct governance and media landscape in Flanders, compared to the other governments involved in the federalized system of Belgium with its federal level and three regions, which we believed would provide more nuanced insights into how local public sentiment affects agency operations and reputation within a more controlled and homogeneous environment. Focusing on the federalized system of Belgium, with its federal level and three regions would provide for a larger sample indeed, but would pose much problems on internal diversity. The federal state and the regional administrations have had the liberty since 30 years to make their own rules and policies regarding agencification and other administrative structures, leading to a different regulatory and governance approach between these different governments. Also, the media landscape functions partly different in the three regions. Focusing on solely the federal (national) level would not help to solve the problem of the small sample either, as there is relatively less agencification at the federal level compared to the Flemish level, where agencification was more pursued as deliberate policy.

### **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 and 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 and 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 and 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, with death being marked by the simultaneous elimination of both structure and function of the involved public organization (Rolland and Roness, 2011). Focusing on maintenance effects provides more depth in the study of structural changes, which is often focusing on birth and death events (MacCarthaigh and Roness 2012). 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 track the structural reforms the organizations experienced during their lifetime. We create a dummy variable that takes the value of 1 if the organization experiences a reform in a certain year and 0 otherwise.

### Main independent variable: media reputation

Scholars have taken different approaches to measuring reputation (Bustos 2021). The present study measures sentiment towards organizations in the traditional (written) news media (Gilad, Maor, and Bloom 2015; Peci 2021; Salomonsen, Boye, and Boon 2021). Though reputation has also been measured through more direct measures of audience perceptions (e.g. Capelos et al. 2016; Overman, Busuioc, and Wood 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 (Deephhouse 2016; Salomonsen, Boye, and Boon 2021; Strömbäck and 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, Shaw, and Weaver 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 and Coleman 2009). In this study, we focus on the sentiment of media coverage 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 supervised machine learning (SML) and natural language processing (NLP) methods to gather and code sentiment to public agencies in news media articles. All content of three national newspapers in Flanders - *Het Laatste Nieuws* (best sold popular newspaper), *De Standaard* (best sold quality newspaper, center-right orientation), and *De Morgen* (second-best sold quality journal, center-left orientation) were scraped from an online newspaper archive.<sup>2</sup> We then used regular expressions to search and filter on articles that contained the name or abbreviation of each of the public organizations under study for the period 2000-2015. Agencies were selected that attracted sufficient media attention, in order to create meaningful indices of media reputation over time. For each article, the sentence(s) in which the organization was mentioned was taken as unit of analysis to calculate

---

<sup>2</sup> <https://academic.gopress.be/>. Formal permission to scrape the archives was granted.

sentiment. Each newspaper article can contain multiple sentences (or: ‘texts’), as an organization can be mentioned multiple times in an article. Similarly, texts can contain the names of multiple organizations, in which case they appear multiple times in the dataset (once for each organization). Table 2 shows the overview of the included organizations and their media attention.

[Insert Table 2 here]

We then followed the procedure as outlined as outlined in Anastasopoulos and Whitford (2018):

*Step 1* – for a subset of texts, perform *human coding* (or: annotation) of sentiment for each public organization;

According to Anastasopoulos and Whitford (2018), models on data produced by expert coders perform significantly better than data produced by nonexperts. Sentiment *towards the studied organizations* in newspaper articles was hand-coded by a team of experts. In the manual coding, 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.

Critically, our measure goes beyond merely capturing sentiment in texts that surround agencies’ names. Rather, the classifiers are trained to recognize *attributions of responsibility*. Therefore, this approach complements theoretical expectations on the role of such attributions, while increasing the likelihood that negative (or positive) frames are effectively targeted at the organization in question.

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. In total, 4916 texts were coded (1533 with negative sentiment; 1327 with positive sentiment; 2056 with neutral sentiment).

*Step 2* – randomly partition human-coded sample into *training subset and a testing subset*, of which the first is used to teach the algorithm the word patterns that belong to each sentiment category, and the second to assess the performance of the algorithm on unseen data;

The BERT (Bidirectional Encoder Representations from Transformers) language model was used to train and test the classifiers. Compared to older models (such as dictionary-based models or word embeddings), the main innovation of BERT is that it can better and more dynamically capture context. This because the representation of a word can change based on the other words in the sentence, allowing BERT to capture the polysemous nature of language more effectively. Furthermore, BERT's transformer architecture allows it to consider the full context of a word by looking at the words that come before and after it. This bidirectional context is a significant leap forward, providing a more comprehensive understanding of sentence structure and meaning (Jurafsky and Martin 2024).

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 Figure 1). In the confusion matrix, each cell allows to compare the instances of predicted labels for a particular class with the actual labels. From the confusion matrix, the precision, recall and f1-score of each model can be calculated (see Table 3). The best performing model had a macro average f1-score of 86.4%, which is well above minimum thresholds of around 65%-70% (Lemmens et al. 2021; Orellana and Bisgin 2023).

[Insert 'Figure 1: Confusion matrix' and Table 3 here]

*Step 3* – apply the validated algorithm to *predict* the scores on the texts in the full dataset.

After the optimal algorithm was selected, a 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 2 shows the percentages of neutral, positive and negative texts for each organization.

On the basis of these data, we calculate the average negative, positive en neutral reputation (variable names ‘Negative reputation’, ‘Positive reputation’ and ‘Neutral reputation’) for each organization by determining the percentage of negative, positive *or* neutral articles in relation to all articles (negative, positive *and* neutral) published about the organization for the three previous years. This involves counting the number of negative (or positive) articles and dividing it by the total number of articles each year, then averaging these three yearly percentages. For instance, when we refer to “more negative” reputations, this means that the relative share of negative articles is higher compared to other organizations and/or time periods. We take a three-year perspective to consider reputation in order to account for the time it takes for reputation to have an effect on reforms. Reputation is a cumulative measure that develops over time, and its influence on driving reforms may require an

extended period to manifest. Additionally, relying on a longer time period of three years helps to avoid drawing conclusions based on isolated or short-term fluctuations, providing a more comprehensive understanding of the organization's overall reputation and its potential impact on reforms.

In Figure 2, an overview of the percentages of positive, negative, and neutral sentiments for each organization over time is visualized. Most organizations predominantly exhibit neutral sentiment, with relatively smaller percentages of positive and negative reputations. However, there are notable fluctuations in sentiment for some organizations. It is important to note that these values represent percentages; thus, for organizations with smaller total numbers, small deviations can result in significant percentage changes.

[Insert 'Figure 2: Evolution of sentiment by organization' here]

### Control variables

We include '*Political turnover*', which reflects whether a new government was instated in the previous year. 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 and Lewis 2004).

### Descriptive statistics

[Insert Tables 4 and 5 here]

Table 4 shows summary statistics for the main variables, while in Table 5 the correlation matrix is presented. Table 4 indicates that, on average, 14% of our observations experienced a reform in a given year. Additionally, we also test for multicollinearity using the variance inflation factor. The mean VIF equals 1.06. The highest VIFs exist for Type (1.13). These values indicate that no collinearity exists between the variables.



## Method and results

To explore the relationship between agencies' reputations and their likelihood of undergoing reform, we employ both logit and probit models on our panel data, with 'experiencing a reform' as the dependent variable. This variable is a binary indicator where '0' represents 'no' and '1' represents 'yes.' We present both models because the probit model allows a random effects approach, which is due to its reliance on a cumulative normal distribution function that complicates the integration of fixed effects. However, the probit model is advantageous in handling cases that consist entirely of either all positive or all negative outcomes. Conversely, the logit model allows for the inclusion of fixed effects but cannot handle such extreme outcome cases, reducing the sample size to 120 observations across 12 organizations. Contrary to the random effects approach of the probit model, the fixed effects approach in the logit model is particularly valuable as it adjusts for time-invariant differences among organizations, isolating the impact of variables that change over time and controlling for all constant unobserved heterogeneity across entities.

Despite the importance of control variables such as structural features, task features, and the availability of alternate service providers, our limited dataset does not allow their inclusion. However, the fixed effects model compensates by implicitly accounting for these variables, though it does not permit a direct examination of their individual impacts. Each iteration of the logit model calculates odds ratios to gauge the strength and direction of the association between reputational variables and the likelihood of reform, where an odds ratio greater than 1 indicates an increased likelihood, and less than 1 indicates a decreased likelihood. We incorporate political turnover into the logit models, as this variable varies over time. In the probit model, we present the beta coefficients.

Furthermore, all models include the percentage of neutral sentiment. Given the dependency among the positive, negative, and neutral percentages—where their sum must total 100%—it is impractical to include all three simultaneously in a single model. By including the percentage of neutral sentiment, we more accurately assess the relative influence of positive or negative sentiments. Results concerning total media attention are detailed in Table 6, while the effects of negative sentiments are shown in Table 7, and positive sentiments in Table 8.

[Insert Tables 6, 7, 8 here]

Looking at the impact of media attention, Table 6 includes two models both the random effects probit model and a fixed effects logit model. Based on both models it is clear that an increase in media attention appears to have no effect on the reform probability.

Looking at the impact of a negative reputation, Table 7 presents two specific setup ups. Column (1) and (3) look at the linear effect of negative reputations on the likelihood to experience reforms. Columns (2) and (4) are added to shed light on possible non-linear effects. As discussed in the theoretical section, negative reputations may be normal facts of life for many public agencies. As negative reputation become fixed attributes of an agencies' reputation, this negativity may no longer elicit much response. By including a squared term of 'Negative reputation', columns (2 & 4) allows to explore such non-linear effects.

The results in column (1 & 3) show that there is no linear effect of negative reputations on the likelihood that agencies experienced reforms. Agencies also have a higher likelihood to be reformed in the years following a switch in government (lag of political turnover). We observe no significant effects of type. The likelihood ratio test indicates a significantly better fit for both columns (2 & 4) which includes the squared term compared to the model without it. Upon studying the models with the squared term, we observe a positive effect for the base term while the squared term exhibits a negative effect. This suggests a non-linear relationship for the effect of a negative reputation, specifically following an inverted U-shaped pattern. To examine this relationship more precisely, we calculate marginal effects based on the probit results<sup>3</sup> and present them in Figure 3. Figure 3 shows that a negative reputation initially increases the likelihood of experiencing a reform. However, as the reputation becomes increasingly negative, the likelihood starts to decrease again. This indicates that the effect of a negative reputation is significant up to a certain point, after which it diminishes. Again, we observe the significant negative effect of political turnover (lag).

[Insert 'Figure 3: Predictive margins' here]

Turning next to the impact of positive reputations (Table 8), we again distinguish between models (1 & 3) that tests the linear impact of positive reputation, and models (2 & 4) that explores potential nonlinear effects through the inclusion of a squared term of positive reputation. Concerning this latter possibility, however, we see that the inclusion of the squared term does not lead to a model improvement. We therefore focus on model 1 & 3, which shows no significant effect of positive reputations on the reform likelihood of agencies.

---

<sup>3</sup> We opt to visualize the squared term effects using the probit re model rather than a logit fe model due to the complexities associated with the fixed effects logistic regression. The fixed effects approach in logit fe removes entity-specific intercepts by differencing, which can obscure the interpretation and visualization of non-linear relationships, such as those involving squared terms. On the other hand, the probit re model, which incorporates random effects, allows us to retain the individual-specific variability and more clearly depict the impact of squared terms on the probability of the outcome, making it a more suitable choice for graphical representation of non-linear dynamics.

## Discussion

The aim of this study was to shed light on the relation between the media reputation of agencies and the likelihood that they experience structural reforms through time. Our findings sharpen extant theoretical insights on the reputation-reform nexus in several ways.

First, our results nuance previous findings on the importance of media attention (in absolute terms) for structural reforms. Agencies that receive more media attention do not succeed in avoiding structural reforms (not supporting H1). Differences in research design might explain divergences from earlier results (Chen, Christensen, and Ma 2023; Bertelli and Sinclair 2015). Perhaps our more general measure of structural reforms (as opposed to terminations) brought in certain types of reform that were not affected by media attention. Also, while we purposely selected media salient agencies in order to generate meaningful indices of media reputation, this design choice might have produced a sample with insufficient variation in media attention to produce effects.

For the sample under study, our results point at another factor: to understand when media attention triggers reforms for media-salient agencies, one needs to consider the tone of coverage. Following studies had confirmed the importance of media attention more generally (Bertelli and Sinclair 2015), as well as the relevance of a dimensional reputational lens (Chen, Christensen, and Ma 2023), this study is the first to theorize and test how an affect-based measure of reputation – accumulated positive and negative sentiment towards public organizations – affects their reform likelihood.

While negative media reputations significantly – and indeed largely positively (cf. H2) – impacted the reform likelihood of public agencies, this effect only manifested in the model that tested for the non-linear effects of negative media reputations. Negative media reputations led to a higher structural reform likelihood, but only up to a certain point. As agencies experienced average reputations of 20% negativity (or more), their likelihood to experience reforms in the following years reduced to zero again. Concerning positive reputations, we observed no significant effect (not supporting H2). The finding that variations in positive reputational signals are irrelevant for agency reforms is surprising in light of literature pointing at agencies' strategic efforts to build and maintain favorable media profiles (Fredriksson, Schillemans, and Pallas 2015), as intangible assets to maximize their bureaucratic autonomy (Carpenter 2001). Yet the finding does support previous studies that argued for the relevance of a negativity bias in understanding human, organizational and political behavior (George et al. 2020; Marvel 2015; Nielsen and Moynihan 2017; Salomonsen, Boye, and Boon 2021).

These findings show that positive and negative are no mirror concepts with equally strong impacts on the reform likelihood of agencies (cf. H3). When it comes to avoiding resource-intensive structural reforms, negative reputations cannot easily be offset by positive coverage.

Our study hints at the existence of distinct affective reputational profiles, with particular effects on the outcome of experiencing reforms. To date, reputation scholars have mainly sought to compare reputational dynamics between agencies on the basis of cognitive dimensions, typically: Carpenter's (2010) four-dimensional framework (e.g. Binderkrantz et al. 2023; Rimkutė 2020). While these studies have led to important insights, our study also points at the existence of distinct affective reputational profiles the implications of which warrant further empirical and theoretical attention.

Our findings inform the two perspectives that have been recognized in the reputation literature on public organizations (Wæraas and Maor 2015). On the one hand, scholars in the political science approach tend to stress the value of reputations as intangible assets that protect agencies from external interventions such as structural reforms (Bustos 2021; Carpenter 2001). In this perspective, reputations are strategically built by agencies to reach positive outcomes such as insulation from structural reforms (which are often politically induced) (Carpenter and Krause 2015). While previous research informs such reputation-building from a dimensional perspective (Chen, Christensen, and Ma 2023) or newspaper audience perspective (Bertelli and Sinclair 2015), our affect-based focus points at the reform-decreasing properties of reputation that are either relatively low or relatively high in negativity.

For those agencies whose average media reputation can be described as predominantly neutral – which we know from previous studies to reflect the majority of agencies across institutional contexts (Boon, Salomonsen, and Verhoest 2019; Deacon and Monk 2001; Schillemans 2012) – keeping negative media attention to a minimum matters in terms of avoiding structural reforms. While the variable neutral reputation in itself was (just) non-significant, this insight complements and enriches Luoma-aho's (2007) insights on the value of neutral reputations. Given that reputation excellence is hard to maintain in a constrained public sector context (see also Wæraas and Byrkjeflot 2012), agencies are better off striving for more realistic levels of reputation and aspiration. There does seem to be a tipping point, however, at which point negative and more heated debate surrounding agencies diminishes their reform likelihood again. One preliminary explanation for this observation is that such crisis settings allow (or force) agencies to enter public debates on their functioning (Boon 2022), more actively garner stakeholder support (Dommett and Skelcher 2014), and showcase their emotional appeal and moral standing (Tom Christensen and Gornitzka 2019).

While rather neutral media profiles seem to benefit from avoiding negative reputational signals, the opposite is true for those agencies where negativity has become so normalized it no longer triggers reform tendencies. In their study on the accumulating effects of reputational history, Salomonsen et al. (2021) use motivated reasoning and attribution theories to explain how audiences such as the media process new incoming information about agencies. At the basis of these theories is the assumption that information mainly triggers our attention when it disconfirms our existing attitudes. Our results may reflect such tendencies: for less negative reputations, incoming information that is negative signals a change in the agency's performance that may require action such as reforms; for (highly) negative media reputations, in contrast, additional negativity does not change the views we already hold, and therefore do not instigate reforms.

On the other hand, the organization science perspective theorizes the role of reputation from an institutionalist and/or social-constructivist perspective (Wæraas and Maor 2015). These scholars, while recognizing the role of the media as an institutional carrier of reputations and agency-related social interactions, tend to place less emphasis on reputation as an independent cause of reforms. This is because reputation is only loosely related to the social and political acceptance – or: legitimacy – of agencies. Therefore, many agencies can operate with negative reputations as a quasi-institutionalized trait of their existence for extended periods of time (Waeraas and Byrkjeflot 2012). Supporting this perspective, our results show that under conditions of more consistent exposure to negative reputational signals, the reform likelihood of agencies drops back to its initial low levels.

While these findings partially support both perspectives, they also invite further reflection and study on how to make sense of these dynamics. From a political science approach to reputation, the observation that the most negative reputations are related to the lowest reform likelihoods is triggering. We outlined the argument from the organizational science perspective to reputation, which questions the close link between reputation and social and political legitimacy. Yet other explanations can be found in theories of blame shifting and blame avoidance (Hood 2013) and public service bargains (Mortensen 2016). Agencification can be driven by beliefs that more autonomy and flexibility will lead to more efficient, innovative, cost-efficient and independent service provision (Van Thiel 2012). Yet scholars have also pointed at the political benefits of agencification, as it may enable blame avoidance by politicians (Flinders 2008). Applied to this study, rather than a model of political responsiveness to reputational signals about the agencies under their remit, we might be witnessing a model of officeholders' acceptance of a situation that allows systematic blame shifting. This would be in line with Hood's (2013) lightning rod argument: even if agencification does not provide the claimed legitimacy improvements (as evidenced in consistently poor reputations), it may help shift blame away from politicians. Our results, therefore, reinforce previous calls to study the interplay of reputational

concerns from different involved actors and organizations in order to understand particular outcomes (Busuioc and Lodge 2016).

Our results also contribute to the literature on structural reforms in government. Explanations for reform occurrence are typically sought in instrumental or institutional perspectives (Kuipers, Yesilkagit, and Carroll 2018), the first stressing (political) agency and the latter stressing structural features. By theorizing and testing the role of reputation, we bring in a novel perspective that may complement and enrich these perspectives, as it sheds light on the nature of information that political-administrative decision-makers rely upon when deciding to engage in reforms (instrumental perspective) or the factors that may interrupt periods of relative reform stability (institutional perspective). Yet a focus on reputation also recognizes a factor that is underappreciated in both perspectives; that is: the independent and strategic role agencies play in crafting a reputation to protect them for political interventions.

## **Conclusion**

This study theorized and tested how negative media reputations of public agencies would impact their reform likelihood. Our results reflect an inverted U-shaped relationship: more negative reputations increase the reform likelihood of agencies, yet up to a certain point at which the reform likelihood drops again. Variations in positive reputational signals, in turn, do not impact the reform likelihood of agencies.

The main contribution of the study is to theorize and test an affect-based measure of reputation, and its impact on agencies' reform experiences. The relation between reputation and high-level political-administrative interventions is widely assumed in the reputation literature (Carpenter 2001), yet has faced little empirical large N scrutiny to date (Chen, Christensen, and Ma 2023; Bustos 2021), none of which focused on the role of sentiment. Furthermore, reform scholars (Kuipers, Yesilkagit, and Carroll 2018; T. Christensen et al. 2007) have dedicated little attention to the role of reputation, even though a reputational perspective has the potential to complement and add to their current explanatory toolkit.

Reflecting the ambition to unravel the role of sentiment, the study relied on novel and state-of-the-art BERT language models to detect sentiment (Jurafsky and Martin 2024). Previous studies have highlighted the challenges of automatically measuring organization-directed sentiment in a comparative and longitudinal set-up notwithstanding (Anastasopoulos and Whitford 2018). This study trained classifiers to effectively recognize attributions of responsibility for positive and negative outcomes. In doing so, this study went beyond mere identifications of positive, negative and neutral

texts surrounding the names of agencies in coverage, thus offering the most refined measure to date of perceived agency performance (or: reputation) in press coverage with what is currently technically feasible.

Our study comes with limitations. First, we recognize the issue of endogeneity. Despite having used fixed effects and having included important control variables, we cannot rule out that omitted variables – e.g. scandals, poor performance – caused a rise in both negative reputation and reform likelihood. Second, and related, this study's focus was not on explaining the contents or causes of reputations. Previous reputation studies related the termination likelihood of agencies to a cognitive and multidimensional measure of reputation that shed light on the performative, technical, moral or procedural contents of reputations (Chen, Christensen, and Ma 2023). These studies related structural reforms – or other outcomes – to *what* people think about agencies, whereas we added to this knowledge base by considering the role of emotions in terms of accumulated sentiment. In doing so, we partly moved away from the reputation literature and its cognitive focus (Maor 2016; Capelos et al. 2016), and relied on insights from political agenda-setting theory that has long recognized the role of sentiment as (politically-perceived) carriers of public legitimacy that urge political-administrative decision-makers towards action to repair an accumulated sense of negativity surrounding an organization. Our limited sample size limited us from training algorithms to detect both sentiment and multiple cognitive dimensions in a meaningful way (that is, if it can be done at all - cf. Carpenter 2020). Future research, however, can take a next step by focusing on a larger – but more domain- and task-specific (Carpenter 2020) – sample of agencies to explore how affective and cognitive perspectives on reputation combine to increase agencies' likelihood to experience reforms. Third, we measured and analyzed "positive reputation" and "negative reputation" as separate variables, rather than attributes that map into an overall reputation. While this approach allowed us to meaningfully compare the distinct impact of more positive versus more negative reputations across organizations on their reform likelihoods (controlling for neutral reputation), future studies – more interested in the nuanced interplay between positive and negative coverage – can follow a different operationalization.

Fourth, the Flemish context is comparable to nation states in the sense of having its own political-administrative system for which it is fully competent for the domains under its remit. However, it also has a particular (Napoleonic) administrative tradition, characterized by high levels of power distance, uncertainty avoidance and a state centric conception of governance with an emphasis on hierarchic control and administrative law. Such traditions, also found in other Latin countries such as France, Italy and Spain, may be particularly conducive to blame shifting effects that may explain the observed effects (Verhoest et al., 2012).

Our findings not only have theoretical implications, they also bring insights that are relevant for practitioners. Media reputations matter, yet practitioners need to be aware of the particularities of the reputational history that characterizes their agency. For agencies with on average neutral media reputations, the chances of experiencing a reform increase substantially when their media profile becomes more negative. In these cases, incoming reputational signals that are negative disconfirm existing patterns of reputational judgment and may trigger reforms. In contrast, agencies that face media criticism on a regular basis are little affected by additional critique when it comes to their reform likelihood. Here, incoming negative reputational signals conform to an accumulated level of criticism that seems to no longer trigger political-administrative need for reform. For these agencies, their legitimacy likely results – and needs to be cultivated – in avenues other than the media.

## Data availability

The data underlying this article are subject to an embargo that runs till the end of the project from which the data originates. The data will be made available on 1/01/2027 at this link: <https://figshare.com/account/home#/projects/221833>.

## Acknowledgments

This work is funded by the Flemish Research Fund (FWO) [grant agreement No. G042523N]. This work was supported by the HORIZON ERC 2022 Advanced Grant [grant agreement No. 101097835] (RegTrust) administered by the European Research Council Executive Agency (ERCEA), and the Flemish Research Fund (FWO) SBO project [grant agreement No. S004124N] and the FWO fundamental project [grant agreement No. G085719N]. This research has benefitted from discussions within the GOVTRUST Centre of excellence. This document reflects only the view of the authors. None of the funding agencies are responsible for any use that may be made of the information it contains.

## References

- Alink, Fleur, Arjen Boin, and Paul T'Hart. 2001. "Institutional Crises and Reforms in Policy Sectors: The Case of Asylum Policy in Europe." *Journal of European Public Policy* 8 (2). Routledge: 286–306. doi:10.1080/13501760151146487.
- Anastasopoulos, L. Jason, and Andrew B. Whitford. 2018. "Machine Learning for Public Administration Research, With Application to Organizational Reputation." *Journal of Public Administration Research and Theory*. doi:10.1093/jopart/muy060.



- Barnes, Carolyn Y., and Julia R. Henly. 2018. "‘They Are Underpaid and Understaffed’: How Clients Interpret Encounters with Street-Level Bureaucrats." *Journal of Public Administration Research and Theory* 28 (2). Oxford Academic: 165–81. doi:10.1093/jopart/muy008.
- Bertelli, Anthony M., and J. Andrew Sinclair. 2015. "Mass Administrative Reorganization, Media Attention, and the Paradox of Information." *Public Administration Review* 75 (6): 855–66. doi:10.1111/puar.12396.
- Binderkrantz, Anne Skorkjær, Jens Blom-Hansen, Martin Bækgaard, Moritz Müller, and Søren Serritzlew. 2023. "The Core of Organisational Reputation: Taking Multidimensionality, Audience Multiplicity, and Agency Subunits Seriously." *Journal of European Public Policy* 0 (0). Routledge: 1–28. doi:10.1080/13501763.2023.2188081.
- Boon, Jan. 2022. "Reputation Management as an Interplay of Structure and Agency: A Strategic-Relational Approach." *Regulation & Governance*. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY. doi:10.1111/rego.12506.
- Boon, Jan, Heidi H. Salomonsen, and Koen Verhoest. 2021. "A Reputation for What, to Whom, and in Which Task Environment: A Commentary." *Regulation & Governance* 15 (2): 428–41. doi:https://doi.org/10.1111/rego.12290.
- Boon, Jan, Heidi Houlberg Salomonsen, and Koen Verhoest. 2019. "The Effects of Organisational Features on Media Attention for Public Organisations." *Policy & Politics* 47 (2): 245–64. doi:10.1332/030557318X15407316633243.
- Boon, Jan, Heidi Salomonsen, Koen Verhoest, and Mette Østergaard Pedersen. 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*, 171–92. New York: Springer.
- Boon, Jan, Jan Wynen, Walter Daelemans, Jens Lemmens, and Koen Verhoest. 2023. "Agencies on the Parliamentary Radar: Exploring the Relations between Media Attention and Parliamentary Attention for Public Agencies Using Machine Learning Methods." *Public Administration*. doi:10.1111/padm.12963.
- Bovens, Mark, and Paul 't Hart. 2016. "Revisiting the Study of Policy Failures." *Journal of European Public Policy* 23 (5): 653–66. doi:10.1080/13501763.2015.1127273.
- Bustos, Edgar O. 2021. "Organizational Reputation in the Public Administration: A Systematic Literature Review." *Public Administration Review* 81 (4): 731–51. doi:https://doi.org/10.1111/puar.13363.
- Busuioac, E.M., and Martin Lodge. 2016. "The Reputational Basis of Public Accountability." *Governance* 29 (2): 247–63. doi:10.1111/gove.12161.
- Capelos, Tereza, Colin Provost, Maria Parouti, Julie Barnett, Jonathan Chenoweth, Chris Fife-Schaw, and Tanika Kelay. 2016. "Ingredients of Institutional Reputations and Citizen Engagement with Regulators." *Regulation & Governance* 10 (4): 350–67. doi:10.1111/rego.12097.
- Carpenter, Daniel. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies 1862-1928*. Princeton: Princeton University Press.
- . 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval." *American Journal of Political Science* 46 (3): 490–505.
- . 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton, NJ: Princeton Univ. Press.
- . 2020. "On Categories and the Countability of Things Bureaucratic: Turning From Wilson (Back) to Interpretation." *Perspectives on Public Management and Governance* 3 (2): 83–93. doi:10.1093/ppmgov/gvz025.
- Carpenter, Daniel, and George A. Krause. 2012. "Reputation and Public Administration." *Public Administration Review* 72 (1): 26–32.
- . 2015. "Transactional Authority and Bureaucratic Politics." *Journal of Public Administration Research and Theory* 25 (1): 5–25. doi:10.1093/jopart/muu012.
- Carpenter, Daniel, and David E. Lewis. 2004. "Political Learning from Rare Events: Poisson Inference, Fiscal Constraints, and the Lifetime of Bureaus." *Political Analysis* 12: 201–32.

- Chen, Sicheng, Tom Christensen, and Liang Ma. 2019. "Competing for Father's Love? The Politics of Central Government Agency Termination in China." *Governance* 32 (4): 761–77. doi:10.1111/gove.12405.
- . 2023. "Reputation Management and Administrative Reorganization: How Different Media Reputation Dimensions Matter for Agency Termination." *Journal of Public Administration Research and Theory* 33 (2): 217–31. doi:10.1093/jopart/muac028.
- Christensen, Jørgen Grønnegård. 1997. "Interpreting Administrative Change: Bureaucratic Self-Interest and Institutional Inheritance in Government." *Governance* 10 (2): 143–74. doi:10.1111/0952-1895.351997035.
- Christensen, T., Per Laegreid, Paul G. Roness, and K.A. Røvik. 2007. *Organization Theory and the Public Sector: Instrument, Culture and Myth*. Routledge.
- Christensen, Tom, and Åse Gornitzka. 2019. "Reputation Management in Public Agencies: The Relevance of Time, Sector, Audience, and Tasks." *Administration & Society* 51 (6). SAGE Publications Inc: 885–914. doi:10.1177/0095399718771387.
- Deacon, David, and Wesley Monk. 2001. "'New Managerialism' in the News: Media Coverage of Quango's in Britain." *Journal of Public Affairs* 1 (2): 153–66.
- Deephouse, David L. 2016. "Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories." *Journal of Management*, June. Sage PublicationsSage CA: Thousand Oaks, CA. doi:10.1177/014920630002600602.
- Dommett, Katharine, and Chris Skelcher. 2014. "Opening the Black Box of Administrative Reform: A Strategic-Relational Analysis of Agency Responses to Termination Threats." *International Public Management Journal* 17 (4): 540–63. doi:10.1080/10967494.2014.958805.
- Flinders, Matthew. 2008. *Delegated Governance and the British State: Walking without Order*. Illustrated edition. Oxford ; New York: Oxford University Press.
- Fredriksson, Magnus, Thomas Schillemans, and Josef Pallas. 2015. "Determinants of Organizational Mediatization: An Analysis of the Adaptation of Swedish Government Agencies to News Media." *Public Administration* 93 (4): 1049–67. doi:10.1111/padm.12184.
- George, Bert, Martin Baekgaard, Adelién Decramer, Mieke Audenaert, and Stijn Goeminne. 2020. "Institutional Isomorphism, Negativity Bias and Performance Information Use by Politicians: A Survey Experiment." *Public Administration* 98 (1): 14–28. doi:10.1111/padm.12390.
- Gilad, Sharon, Moshe Maor, and Pazit Ben-Nun Bloom. 2015. "Organizational Reputation, the Content of Public Allegations, and Regulatory Communication." *Journal of Public Administration Research and Theory* 25 (2): 451–78. doi:10.1093/jopart/mut041.
- Helfer, Luzia, and Peter Van Aelst. 2020. "Why Politicians React to Media Coverage: A Comparative Experiment of Political Agenda-Setting." *The Agenda Setting Journal* 4 (1): 88–108.
- Hollibaugh, Gary E. 2019. "The Use of Text as Data Methods in Public Administration: A Review and an Application to Agency Priorities." *Journal of Public Administration Research and Theory* 29 (3): 474–90. doi:10.1093/jopart/muy045.
- Hood, Christopher. 2013. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton University Press. <http://press.princeton.edu/titles/9353.html>.
- Hvidman, Ulrik, and Simon Calmar Andersen. 2016. "Perceptions of Public and Private Performance: Evidence from a Survey Experiment." *Public Administration Review* 76 (1): 111–20. doi:10.1111/puar.12441.
- Jacobs, Sandra, and Thomas Schillemans. 2015. "Media and Public Accountability: Typology and Exploration." *Policy & Politics* 44 (1): 23–40.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2017. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.
- Jonkman, Jeroen G. F., Mark Boukes, Rens Vliegthart, and Piet Verhoeven. 2020. "Buffering Negative News: Individual-Level Effects of Company Visibility, Tone, and Pre-Existing Attitudes on Corporate Reputation." *Mass Communication and Society* 23 (2): 272–96. doi:10.1080/15205436.2019.1694155.

- Jurafsky, Daniel, and James Martin. 2024. *Speech and Language Processing (Third Edition)*. Upper Saddle River, N.J.: Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3/>.
- Kleizen, Bjorn, Koen Verhoest, and Jan Wynen. 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): 349–67. doi:10.1111/padm.12399.
- Kuipers, Sanneke, Kutsal Yesilkagit, and Brendan Carroll. 2018. "Coming to Terms with Termination of Public Organizations." *Public Organization Review* 18 (2): 263–78. doi:10.1007/s11115-017-0376-4.
- Lemmens, Jens, Tess Dejaeghere, Tim Kreutz, Jens Van Nooten, Ilia Markov, and Walter Daelemans. 2021. "Vaccinpraat: Monitoring Vaccine Skepticism in Dutch Twitter and Facebook Comments." *Computational Linguistics in the Netherlands Journal* 11 (December): 173–88.
- Lewis, David E. 2002. "The Politics of Agency Termination: Confronting the Myth of Agency Immortality." *The Journal of Politics* 64 (1): 89–107.
- Luoma-aho, Vilma L. 2007. "Neutral Reputation and Public Sector Organizations." *Corporate Reputation Review* 10 (2). doi:10.1057/palgrave.crr.1550043.
- MacCarthaigh, Muiris, and Paul G. Roness. 2012. "Analyzing Longitudinal Continuity and Change in Public Sector Organizations." *International Journal of Public Administration* 35 (12): 773–82. doi:10.1080/01900692.2012.715567.
- MacCarthaigh, Muiris, Paul G. Roness, and Külli Sarapuu. 2012. "Mapping Public Sector Organizations: An Agenda for Future Research." *International Journal of Public Administration* 35 (12): 844–51. doi:10.1080/01900692.2012.715568.
- Maor, Moshe. 2016. "Missing Areas in the Bureaucratic Reputation Framework." *Politics and Governance* 4 (2): 80–90. doi:10.17645/pag.v4i2.570.
- Maor, Moshe, Sharon Gilad, and Pazit Ben-Nun Bloom. 2013. "Organizational Reputation, Regulatory Talk, and Strategic Silence." *Journal of Public Administration Research and Theory* 23 (3): 581–608. doi:10.1093/jopart/mus047.
- Marvel, John 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): 209–27. doi:10.1080/10967494.2014.996627.
- McCombs, Maxwell E., Donald L. Shaw, and David H. Weaver. 2014. "New Directions in Agenda-Setting Theory and Research." *Mass Communication and Society* 17 (6). Routledge: 781–802. doi:10.1080/15205436.2014.964871.
- Mortensen, Peter B. 2016. "Agencification and Blame Shifting: Evaluating a Neglected Side of Public Sector Reforms." *Public Administration* 94 (3): 630–46. doi:10.1111/padm.12243.
- Müller, A.C., and S. Guido. 2018. *Introduction to Machine Learning With Python*. O'Reilly.
- Nielsen, Poul A., and Donald P. Moynihan. 2017. "How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy." *Journal of Public Administration Research and Theory* 27 (2): 269–83. doi:10.1093/jopart/muw060.
- Orellana, Salomon, and Halil Bisgin. 2023. "Using Natural Language Processing to Analyze Political Party Manifestos from New Zealand." *Information* 14 (3). Multidisciplinary Digital Publishing Institute: 152. doi:10.3390/info14030152.
- Overman, Sjors, Madalina Busuioc, and Matthew Wood. 2020. "A Multidimensional Reputation Barometer for Public Agencies: A Validated Instrument." *Public Administration Review* n/a (n/a). doi:10.1111/puar.13158.
- Park, Sanghee. 2013. "What Causes the Death of Public Sector Organizations? Understanding Structural Changes and Continuities in Korean Quangos." *International Public Management Journal* 16 (3). Routledge: 413–37. doi:10.1080/10967494.2013.825491.
- Peci, Alketa. 2021. "Agencies in the News? Public Agencies' Media Evaluations in a Low-Trust Context." *Governance*. doi:10.1111/gove.12579.

- Rimkutė, Dovilė. 2020. "Building Organizational Reputation in the European Regulatory State: An Analysis of EU Agencies' Communications." *Governance* 33 (2): 385–406. doi:<https://doi.org/10.1111/gove.12438>.
- Rourke, Francis E. 1984. *Bureaucracy, Politics, and Public Policy*. Boston, MA: Little Brown.
- Salomonsen, Heidi Houlberg, Stefan Boye, and Jan Boon. 2021. "Caught up or Protected by the Past? How Reputational Histories Matter for Agencies' Media Reputations." *Journal of Public Administration Research and Theory* 31 (3): 506–22. doi:10.1093/jopart/muaa056.
- Schattschneider, E.E. 1960. *The Semi-Sovereign People*. New York: Holt.
- Scheufele, Dietram A. 2000. "Agenda-Setting, Priming, and Framing Revisited: Another Look at Cognitive Effects of Political Communication." *Mass Communication and Society* 3 (2–3). Routledge: 297–316. doi:10.1207/S15327825MCS0323\_07.
- Schillemans, Thomas. 2012. *Mediatization of Public Services: How Organizations Adapt to News Media*. Elsen: Peter Lang.
- Strömbäck, Jesper. 2008. "Four Phases of Mediatization: An Analysis of the Mediatization of Politics." *The International Journal of Press/Politics* 13: 228–46.
- Strömbäck, Jesper, and Frank Esser. 2014. "Mediatization of Politics: Towards a Theoretical Framework." In *Mediatization of Politics: Understanding the Transformation of Western Democracies*, 3–28.
- Thesen, Gunnar. 2013. "When Good News Is Scarce and Bad News Is Good: Government Responsibilities and Opposition Possibilities in Political Agenda-Setting." *European Journal of Political Research* 52 (3): 364–89. doi:10.1111/j.1475-6765.2012.02075.x.
- Van Dooren, Wouter, Geert Bouckaert, and John Halligan. 2015. *Performance Management in the Public Sector (2nd Edition)*. Routledge.
- Van Thiel, Sandra. 2012. "Comparing Agencies across Countries." In *Government Agencies: Practices and Lessons from 30 Countries*, 18–29. Palgrave Macmillan.
- Verhoest, Koen, Sandra Van Thiel, Geert Bouckaert, and Per Laegreid. 2012. *Government Agencies: Practices and Lessons in 30 Countries*. London: Palgrave Macmillan.
- Vliegenthart, Rens, Stefaan Walgrave, Frank R. Baumgartner, Shaun Bevan, Christian Breunig, Sylvain Brouard, Laura Chaqués Bonafont, et al. 2016. "Do the Media Set the Parliamentary Agenda? A Comparative Study in Seven Countries." *European Journal of Political Research* 55 (2): 283–301. doi:<https://doi.org/10.1111/1475-6765.12134>.
- Waeraas, Arild, and Haldor Byrkjeflot. 2012. "Public Sector Organizations and Reputation Management: Five Problems." *International Public Management Journal* 15 (2): 186–206.
- Wæraas, Arild, and Moshe Maor, eds. 2015. *Organizational Reputation in the Public Sector*. 1st edition. New York, NY: Routledge.
- Weiner, B. 1985. "An Attributional Theory of Achievement Motivation and Emotion." *Psychological Review* 92 (4). US: American Psychological Association: 548–73. doi:10.1037/0033-295X.92.4.548.
- Wilson, J.Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.
- Wu, H. Denis, and Renita Coleman. 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). SAGE Publications Inc: 775–89. doi:10.1177/107769900908600404.
- Wynen, Jan, Koen Verhoest, and Bjorn Kleizen. 2019. "Are Public Organizations Suffering from Repetitive Change Injury? A Panel Study of the Damaging Effect of Intense Reform Sequences." *Governance* 32 (4): 695–713. doi:10.1111/gove.12404.

## Tables and figures

Table 1: Theoretical model

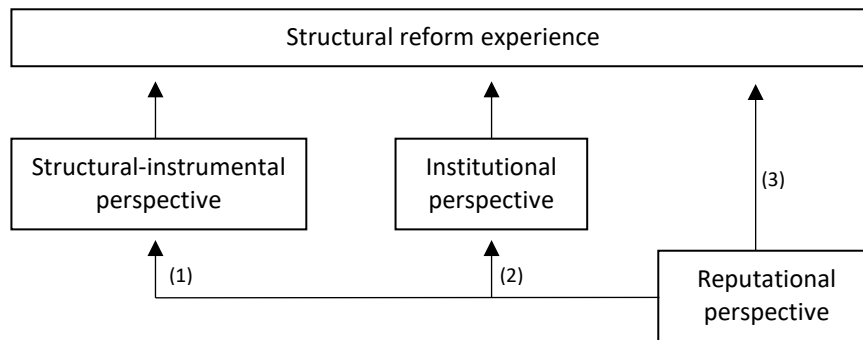


Table 2: Overview of organizations and their media attention

Organization	Number of texts	% neutral	% positive	% negative
AGION	615	78%	14%	8%
AgOndernemen	275	56%	38%	7%
BLOSO	8157	74%	20%	6%
IWT	147	48%	41%	11%
JW	206	61%	25%	15%
KandG	6026	70%	16%	14%
OVAM	3738	80%	13%	7%
VAPH	215	52%	13%	35%
VDAB	11227	68%	23%	9%
VLM	1668	76%	17%	7%
VMM	2943	82%	11%	7%
VMSW	249	70%	17%	13%
VREG	494	81%	8%	11%
ZG	1316	81%	11%	8%
<b>Total / average</b>	<b>37276 (Total)</b>	<b>70% (Avg)</b>	<b>19% (Avg)</b>	<b>11% (Avg)</b>

Table 3: Performance statistics

	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Positive sentiment	87.3	88.6	88.0
Negative sentiment	86.4	86.5	86.5
Neutral sentiment	85.3	84.2	84.7
Macro averaged	86.3	86.4	86.4

Table 4: Descriptive statistics (N=132, across 14 organizations)

Variable	Mean	Sd.
Experiencing a reform	0.14	0.34
Total media attention	167.84	191.35
Average percentage negative reputation (t-1, t-2, t-3)	10.58	8.45
Average percentage positive reputation (t-1, t-2, t-3)	18.93	12.68
Average percentage neutral reputation (t-1, t-2, t-3)	70.49	14.38
Political turnover	0.35	0.48

Table 5: Correlation matrix

Variables		(1)	(2)	(3)	(4)	(5)	(6)
Experiencing a reform	(1)	1					
Total media attention	(2)	0,0223	1				
Average percentage negative reputation (t-1, t-2, t-3)	(3)	-0,054	-0,179	1			
Average percentage positive reputation (t-1, t-2, t-3)	(4)	-0,009	-0,089	-0,118	1		
Average percentage neutral reputation (t-1, t-2, t-3)	(5)	0,0394	0,1837	0,4834*	0,8121*	1	
Political turnover	(6)	-0,088	0,0616	0,1078	-0,019	-0,047	1

\*p<0.05 (5,3: p-value=0.000/ 5,4: p-value=0.000)

Table 6: Regression results ‘Total media attention’

Variables	Probit RE	Logit FE
	$\beta$	OR
Total media attention	0.000460 (0.000368) [0.211]	1.005 (0.00535) [0.390]
Political turnover (lagged)	-0.688** (0.304) [0.024]	0.271** (0.177) [0.045]
Constant	-0.968*** (0.110) [0.000]	
Observations	132	120
Number of groups (org)	14	12

Robust standard errors in parentheses; P-values between square brackets  
 (\*\*\*)P<.01, \*\*P<.05, \*P<.1.)

Table 7 Regression results ‘Negative reputation’

Variables	Probit RE		Logit FE	
	(1)	(2)	(3)	(4)
	$\beta$	$\beta$	OR	OR
Negative reputation	-0.00835 (0.0152) [0.582]	0.547** (0.214) [0.011]	0.880 (0.0889) [0.206]	4.905* (4.128) [0.059]
Negative reputation (sq)		-0.0240** (0.00976) [0.014]		0.911* (0.0440) [0.053]
Neutral reputation	0.00156 (0.00999) [0.876]	0.00464 (0.00841) [0.581]	0.968 (0.0572) [0.586]	0.933 (0.0598) [0.280]
Political turnover (lagged)	-0.680** (0.299) [0.023]	-0.784*** (0.295) [0.008]	0.290* (0.190) [0.059]	0.282* (0.195) [0.067]
Constant	-0.920 (0.796) [0.248]	-3.868*** (1.384) [0.005]		
Observations	132	132	120	120
Number of groups (org)	14	14	12	12
LR test squared term	$\chi^2(1)=7.97***$		$\chi^2(1)=9.68***$	

Robust standard errors in parentheses; P-values between square brackets (\*\*\*)P<.01,  
 \*\*P<.05, \*P<.1.)

Table 8 Regression results ‘Positive reputation’

Variables	Probit RE		Logit FE	
	(1)	(2)	(3)	(4)
	$\beta$	$\beta$	OR	OR
Positive reputation	0.00835 (0.0152) [0.582]	-0.0273 (0.0279) [0.328]	1.136 (0.115) [0.206]	1.033 (0.168) [0.842]
Positive reputation (sq)		0.000576 (0.000406) [0.156]		1.001 (0.00186) [0.450]
Neutral reputation	0.00991 (0.0149) [0.506]	0.00864 (0.0139) [0.534]	1.100 (0.107) [0.328]	1.123 (0.119) [0.270]
Political turnover (lagged)	-0.680** (0.299) [0.023]	-0.662** (0.307) [0.031]	0.290* (0.190) [0.059]	0.304* (0.199) [0.069]
Constant	-1.754 (1.303) [0.178]	-1.290 (1.254) [0.304]		
Observations	132	132	120	120
Number of groups (org)	14	14	12	12
LR test squared term	$\chi^2(1)=0.82$		$\chi^2(1)=0.56$	

Robust standard errors in parentheses; P-values between square brackets (\*\*\*P<.01, \*\*P<.05, \*P<.1.)