Organizational Memory, Reputation and Performance: Differential Performance Response to Reputational Threats*

Moshe Maor  
Wolfson Family Professor of Public Administration  
Department of Political Science  
The Hebrew University of Jerusalem  
Mount Scopus, Jerusalem 91905  
moshe.maor@mail.huji.ac.il

Raanan Sulitzeanu-Kenan  
Senior Lecturer  
Department of Political Science and Federmann School of Public Policy  
The Hebrew University of Jerusalem  
Mount Scopus, Jerusalem 91905  
rasulke@mscc.huji.ac.il

Abstract. How do reputational threats affect the performance of public agencies? Based on a quantitative and qualitative analyses of reputation and outputs data related to the main service delivery agency for the Australian Government in the field of social policy and administration, we find that the effect of negative media coverage on agency performance is moderated by the level of its previous year performance, i.e., negative coverage is followed by an increase in agency outputs when previous year performance is below average, and a decrease in agency outputs when it is above average. The two types of performance response to reputational threats appear to be the result of the agency’s increased interest in change following reputational threats, which is channeled to activities that are internally identified as lagging. Our findings suggest that an agency’s response to reputational threats is endogenously differential, among others, through changes in its organizational outputs or other activities. These findings also empirically demonstrate stable relationships between organizational memory, with reputation management and performance.

* Paper prepared for presentation at the 23rd World Congress of Political Science, July 2014, Montreal. Earlier versions of the paper were presented at the 2014 Accountability Workshop, LSE, the 2014 IRSPM Annual Conference, Ottawa, and the 2014 SOG/IPSA Conference, Jerusalem. The authors thank the Levi Eshkol Institute for Economic, Social, and Political Research at the Hebrew University of Jerusalem, and the Fritz Thyssen Stiftung (Az 10.11.2.132) for their financial support. The theoretical development of this paper was also facilitated by a grant from the Israel Science Foundation. The authors also thank the Australian Government Department of Human Services for its assistance in this project, as well as Alon Burstein for his superb research assistance on this paper.
How do reputational threats affect the performance of public agencies? Assuming that reputational threats affect an agency’s performance, does the agency’s sensitivity to such concerns vary based on the agency outputs? Despite a recent upsurge of research on organizational performance, little attention has been devoted to the understanding of these issues. Agency output is a performance measure which refers to the number of units of goods and services produced by an agency, an increase of which could be credited to the work of the agency. Studies that aim at identifying factors that account for agency performance (e.g., Boyne, 2003; Ashworth, Boyne, and Entwistle, 2010) have not considered the impact of negative agency coverage. However, in a recent assessment of the effects of organizational resources on public agency performance, Lee and Whitford (2010) have found that an agency’s public reputation, among others, has a positive impact on the agency’s performance, measured by the level of goal achievement.

Relatedly, Krause and Douglas (2005) found that reputational maintenance is more important to an agency than succumbing to political pressures and therefore, there is no observable relationship between political insulation and bureaucratic performance. Instead, agencies employ imitative practices in order to avoid being labeled inferior. These findings were corroborated in a related study which suggests that differences in agency design at the federal level do not explain differences in the quality of administrative performance (Krause and Douglas, 2006). Furthermore, Carpenter has argued that audience judgments of an agency may simultaneously relate to its output performance, the expertise of its staff, its values and/or the legality of its processes (Carpenter, 2010). While these studies are insightful, important unanswered questions remain regarding the processes and mechanisms by which reputation maintenance affects
performance. The results of this study offer a nuanced response to this question, by identifying the conditions that determine the different performance responses to reputational threats.

We analyze the relationship between an agency’s reputational threats, as reflected in the printed media, and its performance, as measured across ten organizational outputs and over a range of ten years. We do not assess agency reputation as a resource (Russo and Fouts, 1997), as Lee and Whitford (2012: 15) did, but rather focus on the agency’s response to reputational threats through changes in its organizational outputs. The notion of reputational threat, we believe, is more nuanced than the idea of “managing the external environment” (e.g., Heclo, 1977; Moore, 1995) by relying on public managers’ “political skills”, often mentioned by public management scholars (e.g., Agranoff and McGuire, 2003; Moynihan and Pandey, 2005). Along with recent recognition that “[…] public organizations competed in political reputation markets for the attention and support of the public and politicians” (Lee and Whitford, 2012: 3), we think that the focus on reputational threat can provide meaningful insights into agency response. Specifically, if the concept of ‘threat’ to agency reputation is so crucial, does negative coverage of the agency, whether conveying punishment for past actions or threat as guiding future conduct, actually lead to changes in public agency outputs? This question is at the heart of our study.

The main empirical analysis relies on a panel study of ten different outputs produced by Centerlink — the main service delivery agency for the Australian government in the field of social policy and administration (Halligan, 2008) — as part of its fight against social benefit fraud, over the period 2000-2010. Our main independent
variable is the valence of the agency’s media coverage over this period, gathered through quantitative content analysis of one national and four regional newspapers with the highest circulation in four of the Australian states and territories. Selection of agency outputs was based on a quantitative analysis of survey responses by eighteen senior executives in charge of the fight against welfare fraud at the agency, regarding the visibility, intrusiveness and chances of failure (i.e., risk) of policy tools employed in this fight. Interpretation of the findings was further assisted by eighteen semi-structured interviews with senior executives employed at Centerlink and at the Department of Human Services.

We find that the effect of negative media coverage on agency performance is moderated by the level of its previous year performance, i.e., negative coverage is followed by an increase in agency outputs when previous year performance is below average and a decrease when it is above average. The two types of performance response to reputational threats appear to be the result of the agency’s increased interest in change following reputational threats, which is channeled to activities that are internally identified as lagging. When these include performance, a greater emphasis is placed on increasing this activity. However, when performance is above average, change is channeled to activities other than performance, such as significant investments in public relations, as supported by interview data. Our findings suggest that a public agency’s response to reputational threats is endogenously differentiated, among others, through changes in its organizational outputs, as part of its more comprehensive allocative decisions. These findings also demonstrate stable relationships between organizational

We seek to make three contributions to public management scholarship. First, we intend to link the literatures on the effects of bureaucratic reputation and those dealing with the determinants of agency performance. These two literatures have developed independently despite the fact that agency performance may be measured by “organizational outputs” (Boyne, Entwistle, and Ashworth, 2010: 4) which can easily and immediately be modified following a managerial decision, and can therefore provide insight into an agency’s immediate reaction to a reputational threat. Second, despite the acknowledged importance of public reputation (Lee and Whitford, 2010) as well as the agency’s need to identify and neutralize threats originating in its external environment (Cho and Ringquist, 2010: 59), we can find no research that examines the effect of reputational concerns on organizational outputs in the public sector. We intend to fill this gap by presenting and testing a framework that is useful for research on the potential role of reputational concerns in determining public agency output and hopefully also for various other performance measures. Third, and most importantly, attention to reputational concerns enables us to identify the endogenous nature of the agency’s output response, one that does not depend on regulation, structures, client characteristics, or on the occurrence of some exogenous shock. Instead, the process of differentially responding to negative media coverage (Gilad, Maor, and Ben-Nun Bloom, 2013), given the level of recent agency performance which is stored in the agency’s memory and is retrieved when necessary, focuses our attention on those considerations of the agency itself that make changes in outputs possible. This is perhaps the most important
comparative advantage of employing a reputational framework. The remainder of this paper unfolds as follows. In the second section, we present the analytical framework; in the third we describe the methodology; in the fourth, the empirical findings; and in the fifth, we discuss the theoretical implications of our findings.

**Analytical Framework**

*Reputation and Agency Responsiveness*

The notion of reputational risk is one of the cornerstones of reputation-driven conceptual frameworks in political science nowadays (e.g., Carpenter, 2001, 2002, 2004, 2010a; Carpenter and Krause, 2012; Gilad and Yogev, 2012; Krause and Douglas, 2005; Krause and Corder, 2007; Maor, 2007, 2010, 2011; Moffitt, 2010; Moynihan, 2012; Whitford, 2002, 2003). At the outset, bureaucratic reputation is defined as a set of symbolic beliefs about the unique or separable capacities, intentions, roles, obligations, history, and mission of an organization, that are embedded in a network of multiple audiences (Carpenter, 2010a: 33, 45). This definition centers on the evaluation of the organization’s unique character and activities by multiple audiences. “Reputation uniqueness” according to Carpenter (2001: 5) refers to the demonstration by agencies that they can create solutions (e.g., expertise, efficiency) and provide services (e.g., moral protection) found nowhere else in the polity.

Recent findings have related to the consequences of reputational concerns for the way agencies approve some drugs more quickly than others (Carpenter, 2002) and allocate resources across tasks (Gilad, 2012), for endogenous construction of jurisdictions (Maor, 2010) and observability of decisions and errors (Maor, 2011), and for the duration
of enforcement decisions (Maor and Sulitzeanu-Kenan, 2013). Other studies have demonstrated the extent and ways by which regulatory agencies manage their reputations through the strategic use of communication (Maor, Gilad, and Ben-Nun Bloom, 2013; Gilad, Maor, and Ben-Nun Bloom, 2013). Carpenter (2001, 2010a) has offered some generalized answers to issues regarding reputation and power, and has summed up the contribution of this scholarly literature: “The lesson of this scholarship is that, when trying to account for a regulator’s behavior, look at the audience, and look at the threats” (Carpenter 2010b: 832; italics in original).

Taking this lesson on board, Maor and Sulitzeanu-Kenan (2013) analyzed the impact that reputational concerns have had on ‘time-to-enforcement’ decisions. Focusing on two enforcement divisions within the FDA’s Center of Drug Evaluation and Research, they found that, as media coverage of the FDA’s consumer protection responsibilities becomes more positive, the agency takes enforcement decisions more slowly; in contrast, more critical media coverage leads to quicker action by the FDA. Taking enforcement as a type of performance output, identifying the effect of reputational threats on the pace of enforcement is tantamount to an effect of such concerns on performance (Maor and Sulitzeanu-Kenan 2013: 36). The current study attempts to take a further step, and assess the effect of such concerns more directly on the quantities of goods and services produced by an agency. Assuming that an agency safeguards its unique reputation (e.g., Carpenter, 2001, 2010a; Heimann, 1997; Krause and Douglas, 2005; Maor, 2010, 2011; Quirk, 1980), it is reasonable to expect that agency outputs will tend to increase under media coverage that is negative in tone, as compared to media coverage with a positive tone. This is because such an agency will have an incentive to compensate for the
reputational loss derived from the negative media coverage by increasing agency outputs. This initial theoretical intuition thus suggests that the agency under attack is likely to demonstrate visible and easily understandable evidence for an increase in its outputs. Thus, we hypothesize that:

*Hypothesis I: When faced with negative media coverage, an agency will increase its level of outputs*

*The Notion of Selective and Differential Responsiveness*

The second step taken in this study is to empirically test a more nuanced theory regarding the effect of reputation protection on performance. We propose that the effect of external reputational threats on agency performance is moderated by the agency’s own judgment of its current performance. While organizations are expected to be responsive to external reputational threats, their response is plausibly the product of integrating such external signals with internal judgments regarding their relevance and validity, given other types of information available to the agency or stored in its organizational memory to be retrieved when necessary (Walsh and Ungson, 1991; Weinberger, Te’eni, and Frank, 2008). This integration is likely to moderate the response of the agency to external reputational threats. Maor, Gilad, and Ben-Nun Bloom (2013) have demonstrated that a regulator *selectively* responded in the media to external allegations, given its assessment of their relative threat to its reputation. Specifically, they have shown that a regulatory agency tends to keep silent on issues regarding which it generally enjoys a strong reputation, and on issues that lie outside its distinct jurisdiction, while responding to
opinions about core functional areas in which its reputation is weaker and areas wherein its reputation is still evolving.

If agency budget is an analogue of agency reputation, the aforementioned example ties in nicely with Macmillan’s (2012) finding that the U.S. Department of Justice’s Anti-Trust Division increases enforcement output when their funding is lower than they deem optimal, while engaging in preserving status-quo enforcement levels when their funding is higher than they deem optimal. This explicit tradeoff between financial security and policy goals is undertaken by agencies in order to protect themselves from external pressure brought about by appropriations politics. Gilad, Maor and Ben-Nun Bloom (2013) have also shown that a regulatory agency differentially responded in the media to external allegations by publicly acknowledging problems and shifting blame to others when faced with claims of under-regulation, and by denying allegations of excessive regulation.

In the context of performance, we therefore expect an agency’s performance patterns to be selectively responsive to external reputational threats when the agency’s internal evaluation of its performance is low, and to exhibit less responsiveness when the agency evaluates its performance as high. We employ actual relative performance as a proxy for the agency’s internal evaluation, and hypothesize that:

Hypothesis II: The effect of negative media coverage on agency performance will be moderated by the level of recent agency performance, i.e., the effect will be stronger following periods of below average outputs, and weaker following periods of above average outputs.
Ultimately, we hypothesize that a public agency’s response to reputational threats will be undertaken, among others, through changes in its organizational outputs. Such a finding entails that reputational concerns create a response mechanism that produces institutional outcomes (agency outputs) over and above the content of rules, guidance, procedures, structures, and office holders. Our attention now turns to the methodological section.

**Research Design**

We investigate the effect that reputational considerations have on agency outputs by focusing on a single case which is theoretically illuminating, relying on two novel datasets that have been constructed as part of our research, and contain agency outputs and media coverage over a decade. These datasets allow us to estimate an agency’s strategic response to reputational threats by analyzing systematic changes in its organizational outputs in a panel design.

**The Policy Domain**

The policy domain examined here is the delivery of welfare benefits and services for retirees, the unemployed, families, caregivers, parents, people with disabilities and minorities, while preventing and investigating welfare fraud. Welfare fraud occurs when individuals or organizations acquire social benefits from a government agency or department, either by obtaining payments to which they are not entitled or by withholding funds that they should pay over to the agency or department. This policy domain was selected for three reasons, which also contributed to the increased attention devoted to the issue by the general public, the media and policymakers during the last decade. First, in
recent years there has been a growing recognition of the negative impact that welfare fraud has on the economies of industrialized countries that acknowledge public responsibility for the provision of some social support and spend very substantial amounts on benefit payments (e.g., UK National Audit Office, 2006; UK Treasury & National Audit Office, 2008; UK Revenue & Customs and Department for Work and Pensions, 2010). Second, criminal prosecutions as well as other anti-fraud measures have created an ongoing controversy in welfare countries regarding the legitimacy of the welfare system, the balance between a punitive approach and prevention of fraud before it occurs, and the stigmatization of welfare recipients (e.g., Bradbury, 1988: 26; Prenzler, 2011). Third, the fight against welfare fraud has gained tremendous momentum following the 2008 financial meltdown and the consequent economic crisis. Many industrialized countries have made combating fraud a major issue in their government objectives, expanding existing prevention and detection measures, and introducing new ones (Prenzler, 2011).

Despite the growing recognition in recent years of the damaging impact of welfare fraud, most scholars do not place the fight against social benefits fraud at the center of analysis. There are some notable exceptions: Etzioni (1982) has investigated anti-fraud initiatives based on prevention that is grounded in an understanding of the constituencies involved; Grover (2005) has looked at advertising social security fraud, and Rowlingson et al. (1997) have studied the role of penalties in this context. Moreover, McKeever (1999a) has examined public reaction to benefit fraud in the U.K., and the rationale for legislative changes in the fight against benefit fraud (McKeever 1999b). Walsh and Martson (2010) have studied the conflation of social security fraud and
overpayment. Sandberg (1993) has shown that tolerance for different forms of abuse has generally decreased in Sweden. And Chunn and Gavigan (2004, 2006) have analyzed “welfare cheats,” the place this concept occupies in attacks on the poor, and the theoretical and empirical questions related to regulation, control, and the relationship between them at particular historical moments. These studies have provided descriptive and normative accounts of this policy domain. However, the possible impact reputational concerns may have had on the fight against welfare fraud has remained uncharted territory.

*Australia’s Centrelink Agency*

Australia has provided welfare benefits to vulnerable and disadvantaged citizens since 1909, and during the period under examination (i.e., 2000-2010), almost one-third of Australians received some kind of direct welfare payment (Prenzler, 2011). In 2008-09, for example, Australia’s federal welfare agency Centrelink distributed approximately $75b to 6.8 million recipients across 140 benefit types on behalf of 27 government departments and agencies (*Centerlink*, 2009: 28; Figures are in U.S. dollars). The availability of numerous benefits during the period under investigation has created opportunities for minor and major welfare fraud. Not surprisingly, media coverage of fraud has revolved around major cases of welfare fraud, alerting the public to the anti-fraud work of government agencies and departments and to the costs of fraud (Prenzler, 2011: 4). In addition, media coverage has fuelled popular opinion against welfare cheats (Prenzler, 2011: 2). Still, estimates of the level of fraud have ranged from “[…] a significant number of people […]” (Bajada, 2005: 184), to less than half of one percent of
social security debt (Karvelas, 2008: 4). Beyond estimates, although conviction rates during the period under investigation stood at around 3000 per year, representing 0.04% of Centerlink’s customers, the derived losses of these cases involved approximately $105m per year in gross savings and amounts targeted for recovery (Prenzler, 2012: xii).

The agency under investigation is Centerlink, which was established in 1997 as the main service delivery agency for the Australian government in the field of social policy and administration (Halligan, 2008). In October 2004, Centerlink autonomy was considerably curtailed with the establishment of the Department of Human Services, which has been entrusted with the development of service delivery policy and the provision of access to social, health and other payments and services (Halligan, 2007). Consequently, the agency’s CEO who had served since 1996 resigned, and in 2005-06, new governance arrangements were created. The agency was formally incorporated into the Department of Human Services in 2011 as a result of the Human Services Act 2011, with the department retaining the Centerlink brand name as part of its set of main policy programs.¹ The period under examination covers 2000-2010, thus encompassing most of the period in which Centerlink was an autonomous and semi-autonomous agency.

Public Agency Outputs

Measuring outputs in the public sector is notoriously more complicated than doing so in the private sector, where a price-weighted measure of overall outputs can be relatively easily calculated (Dunleavy and Carrera, 2013). In contrast of the private sector, some public goods and services are provided for free or at a subsidized price, some are mandatory, and others are imposed on citizens (e.g., prisons). Additionally, ethical
considerations bar public institutions from targeting particular segments of the population (e.g., racial profiling), and in many policy domains it is not practical to rely on citizens’ satisfaction as a general measure of outputs and performance. This complexity is deepened when no cross-agency comparison is forthcoming due to the fact that only one agency produces the relevant goods or services in a given polity. For these reasons the recommended practice of measuring public outputs relies on internal measures of social delivery. Such measures need to encompass the various types of outputs for lack of a common currency of measure (Kelly 2005). In line with these considerations, the measure of agency output adopted here relies on the entire set of agency activities, which were identified by senior agency officials interviewed, as critical for achieving the agency’s overall task of monitoring the provision of benefits and fighting welfare fraud. This approach requires two assumptions that should be made explicit. First, that at the level of an individual output, the units of goods and services have a relatively homogeneous quality. And second, that each of the outputs under consideration provides a comparable contribution to the task performed by the agency.

*Centerlink*’s outputs data over the period under investigation have been obtained from both the agency’s annual reports and the agency itself upon our request. Table 1 presents the list of agency outputs which were selected. We focus on the main detection strategies; namely, eligibility reviews, service profiling on the basis of recipients’ characteristics (i.e., a risk-based approach), job capacity assessment and data matching with government and non-government institutions (e.g., universities). We omit the reporting of suspected fraud by the general public because the agency, once establishing the infrastructure (e.g., fraud tip-off line), has played a passive role in employing this
strategy for detection purposes. In addition, we focus on the aggregate measure of fraud investigations. This measure consists of investigations which vary by target (e.g., identity-based and cash-economy investigations), source (e.g., tip-offs), mode (e.g., optical surveillance), and location (e.g., investigations of customers living in Australia or overseas). We also take into account the main agency activities which are derived from investigations, namely, prosecution referrals, formal repayment agreements, and debt recovery by private contractors. In addition, we consider the main measure employed by the agency to increase public awareness regarding the fight against welfare fraud as well as deterrence, namely, public campaigns.

TABLE 1 ABOUT HERE

Data and Operationalization

Our measure of reputational threat rests on a quantitative content analysis of the agency’s media coverage. To conduct our analysis, we made use of a national newspaper – the *Australian* – as well as four daily newspapers — *the Age, the Courier Mail, the Hobart Mercury, and the Sydney Morning Herald*; each is the most widely-circulated newspaper in one of four Australian states. Using the *Lexis-Nexis* database, we identified all articles published between January 1, 2000 and December 31, 2010 that addressed the agency and any of its policy activities. These newspaper articles were filtered in order to exclude letters and obituaries. Furthermore, in order to avoid endogeneity of the press valence measure, the measure of media valence relied selectively on press articles in
which the sources did not include the agency; thus the valence is based solely on external sources. The remaining data consisted of 771 articles which were coded for valence (positive, negative and neutral). In order to assess the reliability of the coding, a random sample of 20 articles was coded independently by three coders. Treating the valence coding as ordinal resulted in acceptable levels of inter-coder reliability (Krippendorff’s α = 0.83). In addition, eighteen semi-structured interviews with senior executives employed at Centerlink and at the Department of Human Services were undertaken during January 2012 as part of this research. All interviews were recorded, fourteen of which were conducted at the agency’s offices, and the remaining four over the phone. Interviewees were selected based on their executive role covering the top three administrative ranks at Centerlink. After each interview, the interviewees were asked to rank the agency’s policy tools (1-10) according to their level of public visibility, intrusiveness and chances of failure.

As noted above, we employ actual recent performance as a proxy for the agency’s internal evaluation of its performance. In order to do so it is important to assess whether this proxy in fact reflects another variable – organization slack – which may account for the hypothesized relationship.³ One may expect that the agency’s responsiveness to external reputational threats would be moderated by organizational slack – “the pool of resources in an organization that is in excess of the minimum necessary to produce a given level of organizational output” (Nohria and Gulati, 1996) – thus such slack is likely to be correlated with the level of output. If this is the case, the hypothesized observed relationship between agency responsiveness and recent performance can actually stem from the availability of slack for performance response, rather than the evaluation by the
agency of its performance. In order to account for the level of slack in the agency we created a measure of slack by dividing the agency’s previous year standardized budget by previous year performance.

To test our hypotheses we conducted a joint analysis of all ten agency outputs, by converting their yearly raw measures into standardized measures. This research design provides a cross-sectional (output) panel (year) dataset. To analyze the data, we estimated the following regression model:

\[ \Delta Y_{ot} = \alpha + \beta_1 V_{t-1} + \beta_2 Y_{ot-1} + \beta_3 V_{t-1} Y_{ot-1} + \beta_4 \frac{\text{budget}_{t-1}}{Y_{ot-1}} + \beta_5 V_{t-1} \frac{\text{budget}_{t-1}}{Y_{ot-1}} + \beta_6 T + \epsilon_{ot} \]

Where \( \Delta Y_{ot} \) represents the change in standardized performance of policy output \( o \) in year \( t \), \( V_{t-1} \) is the measure of one-year lagged press valence (a higher value indicates more positive coverage), \( Y_{ot-1} \) is the one-year lagged performance of the policy output, \( \frac{\text{budget}_{t-1}}{Y_{ot-1}} \) is the measure of one-year lagged slack, \( T \) represents the yearly trend, and \( \epsilon_{ot} \) represents stochastic error. The coefficient (\( \beta_3 \)) of the interaction term \( V_{t-1} Y_{ot-1} \) estimates the difference in the agency’s performance response to negative media coverage over varying levels of recent performance, and thus provides the theoretically central estimation.

Our analysis is based on changes in a set of agency outputs over time. With such data one must assume that ‘regression to the mean’ (RTM) may affect the temporal dynamic of the data, unless shown otherwise (Barnett et al. 2005). This assumption is particularly relevant as our second hypothesis predicts that one-year lagged performance is expected to influence the level of consequent year performance. Following the method suggested by Barnett et al. (2005) we evaluate the magnitude of the RTM effect by plotting a scatterplot of change in the standardized output (follow-up minus one-year
lagged output) against one-year lagged output, presented in Figure 1. The fitted line was obtained by linear regression of the change values on baseline values. A clear pattern of RTM is apparent in the plot, as outputs that follow unusually low levels have tended to increase (so that change values are likely to be above zero), outputs that follow unusually high levels have tended to decrease (so that change values are likely to be below the zero), and average output levels have tended to be followed by no change (on average). This pattern depicts a classic RTM effect, estimated at -0.634 standard deviation change for a +1 standard deviation in the one-year lagged performance. Note that the RTM effect is expected to be uniform for a given level of lagged performance, while our second hypothesis predicts that the change in performance will also depend on media valence by estimating the interaction between lagged performance and media valence; therefore it is not expected to bias our estimations.

------------------------------------------------------------------------------------------------------------

FIGURE 1 ABOUT HERE

------------------------------------------------------------------------------------------------------------

Given that all the outputs are produced by the same agency, we should expect a cross-sectional correlation (e.g., the agency’s finite budget entails that an increase in one output leaves fewer resources for others) (Beck and Katz, 1995, Driscoll and Kraay, 1998). In order to estimate standard errors that are robust to disturbances resulting from heteroscedasticity, autocorrelation, and cross-sectional dependence, we used the Driscoll and Kraay (1998) approach, implemented by the xtscc Stata command (Hoechle 2007). Data are clustered by policy outputs.
Results

Table 2 presents five regression analyses with Driscoll-Kraay standard errors, with change in standardized agency output as dependent variable. Model one provides a baseline model by regressing change in standardized agency outputs on one-year lagged standardized outputs and time.$^4$ Model 1 indicates that one-year lagged output and time are fairly good predictors of consequent year change in output. The former represents a clear empirical implication of RTM, and the latter, the particular overall trend in the data over the research period. This model accounts for 45% of the variance in the change in agency outputs, and provides a baseline for evaluating the contribution of the variables of theoretical importance for predicting agency output in the following models.

Model 2 adds one-year lagged media valence, in order to test the first hypothesis. This analysis indicates no significant association between press valence and agency performance, thus providing no support for the first hypothesis. However, it is possible that the moderating mechanism suggested in our second hypothesis does not allow us to detect a statistically significant main effect of media valence.

In order to test our second hypothesis, Model 3 adds an interaction term between press valence and one-year lagged performance. The fact that the coefficient for this interaction term is statistically significant suggests that the change in agency output predicted by previous year media valence is significantly different when lagged performance is relatively high and low, providing support for our second hypothesis. Models 4 and 5 further assess the robustness of this finding by controlling for the potential effect of slack. Model 4 adds a simple control for slack, and model 5 adds an interaction term between lagged media valence and slack – allowing us to estimate the fit
of the alternative explanation to the data. The size and statistical significance of the ‘valence X lagged-performance’ interaction term remain substantively unchanged across models 4 and 5. Note that this finding holds although model 5 provides support for the expectation that slack also moderates the agency’s performance response to media valence. Interpreting the results of model 5 suggests that when the agency’s recent performance level is average, media valence appears to have no effect on subsequent agency performance ($\beta=.004, p=.992$). When recent performance level is one standard deviation above average, media valence appears to have a positive effect on the change in agency performance ($\beta=.624, p < .01$), and when recent performance is one standard deviation below average, the effect of media valence is negative ($\beta=-.413, p < .01$). The difference between the latter two predictors is statistically significant ($p < .001$).

TABLE 2 ABOUT HERE

These findings provide support for hypothesis two, while introducing another unexpected finding. As predicted by hypothesis two, when recent agency performance is relatively low, the agency appears to increase its outputs the more negative its media coverage is, and this relationship is absent when agency output level is about average. However, when agency performance is relatively high, its performance appears to decrease the more negative its media coverage has been in the preceding year. Figure 2 graphically presents these findings (estimates based on model 5). The x-axis represents three levels of one-year lagged media valence – ‘positive’ (90th percentile of the valence score), neutral (50th percentile) and negative (10th percentile). The y-axis depicts
consequent change in standardized performance. The three colored bars present the relationship between the two dimensions (media valence and performance change) across three levels of lagged agency performance – ‘low’ (-1SD, blue bar), mean (red bar), and ‘high’ (+1SD, green bar), with 95% confidence intervals. While the RTM effect results in the overall vertical direction of each of different bars (in general, positive changes follow low performance, and negative changes follow positive performance), the pattern of change in the size of the bars depicts a varying effect of media valence across the levels of performance. Following positive media coverage, changes in agency performance are minimal, suggesting negligible improvement when performance is low and negligible decline when performance is high, and in both cases the changes are smaller than the mean RTM effect. However, following negative media coverage, changes in agency performance are largest, resulting in strong improvement when performance is low and a large decline when performance is high. These changes are clearly larger than the effects of positive coverage, as well as the mean RTM effect. In the following section we further discuss these findings.

FIGURE 2 ABOUT HERE

Discussion

Fifteen years ago, the topic of bureaucratic reputation was not even a promissory note, let alone a tangible reality. A decade of research makes it clear that reputational concerns influence agency behavior (for a review see Maor, in press). The findings of this study add another layer to the growing interest in this topic, suggesting the agency’s
responsiveness to reputational threats is not automatic, but rather dependent on the agency’s internal evaluation of its performance. When the agency’s performance is below average, negative media coverage leads to increased performance in the following year. However, this relationship is nullified when agency performance is about average, and is reversed when performance is above average – i.e., performance tends to decrease following negative media coverage. While the former findings confirm hypothesis two, the latter was not expected, and an ex-post development of our theoretical framework is suggested.

The results depicted in Figure 2 indicate that change in performance is largest when media coverage is negative and minimal when it is positive. This finding conforms to the notion that agencies are sensitive to external reputational threats, and that this sensitivity is also manifested in their performance. Before moving more closely to the differential response of the agency to such reputational threats, let us consider a broader view of the agency’s response to such threats. The main challenge faced by an agency, as any other organization for that matter, is how to allocate its limited resources to a set of activities – an exercise that determines the particular mix of activities the agency will undertake. In this context the performance responsiveness of an agency is likely to abide by the following logic: When its reputation is challenged (negative coverage) the agency is expected to look for effective ways to respond in order to address its reputational decline. The alternatives available to the agency are given by the possible different mixes of its activities. This process of evaluating alternatives for action and selecting the one to be pursued is undertaken by relying, among others, on performance information stored in the agency’s organizational memory. Thus, when a reputational threat is joined by below
average performance, the agency's performance is likely to take greater importance in the revised resource allocation mix, the observed result of which would be a distinct increase in its performance the following year. However, when a reputational threat is joined by above-average performance, the agency’s drive for change is likely to be channeled to activities other than performance – for example, public relations during natural disasters, community engagements, stakeholder consultation, and so on. Assuming that the agency’s resources are finite, such allocation choices would result in a decline in performance.

For example, during Tropical Cyclone Yasi in North Queensland and bushfires in Western Australia in 2010-11, *Centerlink* mobilized 2/3 of its staff to recovery centers in the disaster area. Although this mobilization led to a 72% fall in prosecution referrals within a year (*Centerlink* 2010-11, 151), it was designed in advance as an attempt to “demonstrate […] the agency’s] service delivery capabilities at its finest” (*Centerlink* 2010-11, 4). *Centrelink* set up mini-mobile offices; its employees were provided with green shirts with the agency’s logo, and the agency used banners on its website in order to build a profile around the presence of its employees on the ground. At the same time, the agency’s communications team gave around 600 media interviews to reinforce the message and the positive attributes of what the agency was doing. The message was that if one needs help, one is welcome to the agency’s recovery centers in the area affected, but one is warned that the agency will do the background checks. So the agency communicated that it was helping those in need but at the same time ensuring that its integrity remained intact. These efforts to advance reputation for effective delivery of crisis payment combined with fraud prevention were evident since 2010. However, in the
early 2000s, when most criticism of Centerlink has revolved around poor service delivery, mostly over its phone service center, the agency tried to build confidence in the integrity of the system. And in the mid-2000s, it advanced reputation for credibility, especially by supplying the media with covert surveillance footage of fraudulent Centerlink customers as well as vision obtained for cash economy operations (e.g., pre-operation surveillance) (Centerlink 2011).8

The theoretical framework advanced here is admittedly distinct from our point of departure in the sense that the agency is responsive to reputational threats, but this responsiveness is not endogenously selective, but rather endogenously differential – i.e., the response varies between types of responses (performance vs. other activities), depending on its self-evaluation, rather than between levels of a particular type of response.

These findings conform to recent findings regarding the effects of reputational threats on the duration of enforcement decisions in another and very different agency (Maor and Sulitzeanu-Kenan, 2013), extending the notion of endogenously selective responsiveness (Maor, Gilad, and Ben-Nun Bloom, 2013), and provides another example of an endogenously differential response (Gilad, Maor, and Ben-Nun Bloom, 2013). These findings also empirically demonstrate stable relationships between organizational memory, with reputation management and performance.

Additional research with other performance measures is required in order to demarcate the extent to which these findings are general. Suitable measures may include judgments of internal or external stakeholders regarding performance (Walker and
Boyne, 2006); organizational inputs and/or outcomes (Cohen and Eimicke, 2008; Linder, 2004), and qualitative performance data (Blasi, 2002).

We still know little about whether changes in public agency outputs have been undertaken across the board, or whether they reflect a strategic decision limited to a carefully selected set of agency outputs. In addition, we still have little knowledge of the relationship between an agency’s reputation and agency behavior over longer time periods as some variation in media coverage may be historically contingent, that is, coming from a short period of time of a decade or so. In addition, this research has dealt solely with one agency, although the findings appear to conform to previous studies on very different agencies. Further research should be devoted to enhancing our understanding of how the reputational considerations of one agency may differ across policy domains and national settings. Although much work needs to be done to unravel how an agency’s basis of reputation impacts upon its behavior, we hope that this study has modestly provided a contribution to the bureaucratic politics literature in this context.
References


Table 1: Centerlink’s Outputs under Examination

<table>
<thead>
<tr>
<th>Agency output</th>
<th>Yearly mean</th>
<th>Period for which data was available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligibility reviews</td>
<td>3,592,860</td>
<td>2000-2010</td>
</tr>
<tr>
<td>Prosecution referrals</td>
<td>4,227.36</td>
<td>2000-2010</td>
</tr>
<tr>
<td>Public campaigns</td>
<td>0.55</td>
<td>2000-2010</td>
</tr>
<tr>
<td>Data matching with government agencies</td>
<td>17</td>
<td>2000-2010</td>
</tr>
<tr>
<td>Number of formal repayment agreements</td>
<td>1,278,419</td>
<td>2000-2010</td>
</tr>
<tr>
<td>Service profile updates</td>
<td>835,571.7</td>
<td>2002-2010</td>
</tr>
<tr>
<td>Data matching with non-governmental institutions</td>
<td>9,889.33</td>
<td>2002-2010</td>
</tr>
<tr>
<td>Investigations into serious non-compliance and fraud</td>
<td>33,350.71</td>
<td>2004-2010</td>
</tr>
<tr>
<td>Job capacity assessments</td>
<td>292,238.2</td>
<td>2006-2010</td>
</tr>
<tr>
<td>Number of mercantile agents used in debt recovery</td>
<td>278,856.8</td>
<td>2006-2010</td>
</tr>
</tbody>
</table>
Table 2: Panel Data Regression Analyses with Driscoll-Kraay Standard Errors – standardized change in agency output as dependent variable

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valence X Lagged-Performance</td>
<td></td>
<td>1.508 (.425)**</td>
<td>1.534 (.424)**</td>
<td>1.586 (.408)**</td>
<td></td>
</tr>
<tr>
<td>Valence X Slack</td>
<td>.057 (.023)*</td>
<td>.056 (.346)</td>
<td>.004 (.345)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagged performance</td>
<td>-.703 (.061)***</td>
<td>-.696 (.066)***</td>
<td>-.977 (.137)***</td>
<td>-.960 (.140)***</td>
<td>-.966 (.143)***</td>
</tr>
<tr>
<td>Year</td>
<td>.215 (.048)**</td>
<td>.224 (.050)**</td>
<td>.186 (.051)**</td>
<td>.149 (.021)***</td>
<td>.142 (.022)***</td>
</tr>
<tr>
<td>Year cube</td>
<td>-.004 (.0006)***</td>
<td>-.004 (.0006)***</td>
<td>-.003 (.0007)***</td>
<td>-.003 (.0004)***</td>
<td>-.003 (.0004)***</td>
</tr>
<tr>
<td>Slack ((\frac{\text{stand} \text{ budget}<em>{t-1}}{y</em>{t-1}}))</td>
<td>-.176 (.115)</td>
<td>-.288 (.206)</td>
<td>-.136 (.148)</td>
<td>.007 (.110)</td>
<td>.041 (.112)</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F)</td>
<td>72.51***</td>
<td>118.12***</td>
<td>35.46***</td>
<td>36.29***</td>
<td>233.02***</td>
</tr>
<tr>
<td>R-squared</td>
<td>.45</td>
<td>.46</td>
<td>.48</td>
<td>.51</td>
<td>.51</td>
</tr>
<tr>
<td>(N)</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>79</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01, ***p < 0.001.
Figure 1: Estimating the magnitude of the 'regression to the mean' (RTM) effect. Mean change in standardized performance as a function of one-year lagged performance suggest that one standard deviation above (below) the mean is followed by average decrease (increase) of 0.634 standard deviation in the consequent year ($p < .001$, $R^2=.24$).
Figure 2: The relationship between lagged media valence and performance change, across lagged performance levels (estimates are based on model 5)
Under the Human Services Legislation Amendment Act 2011, Medicare Australia and Centerlink services were integrated into the Department of Human Services. This major change has led to a transformation of the Department’s structure and organizational culture, as well as to a change in the role of Centerlink (Department of Human Services 2012, 10). This, in turn, has closed the window for extending the period under examination.

The exact search specification was as follows: "((welfare) OR (social security) OR (social benefits) OR (Centrelink)) AND ((fraud) OR (scam) OR (deceit) OR (cheat) OR (overpayment) OR (investigation) OR (plead guilty) OR (found guilty))”.

We are grateful to Donald Moynihan for pointing our attention to this possibility.

Based on assessing several specifications of time-trend suggest a non-linear trend, best described by a cubic relationship.

Senior official, Department of Human Services, Interview with Moshe Maor, Canberra 31.1.2012.

Senior official, Centerlink, Department of Human Services, Interview with Moshe Maor, Canberra 30.1.2012.

Senior official, Department of Human Services, Interview with Moshe Maor, Canberra 31.1.2012.

Senior official, Department of Human Services, Interview with Moshe Maor, Canberra 31.1.2012.