Why Public Employees Manipulate Performance Data

Prosocial Impact, Job Stress, and Red Tape

Alexander Kroll[a] and Dominik Vogel[b]

[a]*Steven J. Green School of International and Public Affairs, Florida International University, Miami, FL, USA;* [b]*Faculty of Business, Economics and Social Sciences, University of Hamburg, Germany*

This is an Accepted Manuscript of an article published by Taylor & Francis in the *International Public Management Journal* on 2021-02-22, available online: http://www.tandfonline.com/10.1080/10967494.2021.1874579.

Why Public Employees Manipulate Performance Data
Prosocial Impact, Job Stress, and Red Tape

While research on the dysfunctional uses of performance data is growing, we are still in search of theories that go beyond system-specific explanations and address data manipulation behavior at the level of the employee. In this article, we conceptualize different gaming responses to performance systems and test a model of performance cheating that emphasizes the critical role of employees' prosocial impact, their job stress, and organizations' red tape. We screen through a sample of almost 10,000 potential subjects and identify 964 public employees who work with performance data. Conducting a list experiment, a technique known to yield unbiased ratings of sensitive behaviors, we find that all three factors tend to reinforce performance cheating among public employees. The article contributes to the extension of causal chains in performance gaming theory via the inclusion of factors that have proven to be influential in behavioral research.

Keywords: performance management, performance gaming, performance cheating, performance information use

Introduction

In Phoenix, Arizona, the veteran affairs health system met a performance target set by the central office in Washington. In order to be able to report that veterans would be able to see a doctor in no more than 14 days, the Phoenix office kept actual waitlists secret and created a manipulated waitlist that showed compliance (Hill and Lynn 2015:323ff.). In Atlanta, Georgia, 35 school administrators had been involved in a conspiracy to inflate student test scores in order to meet performance targets and receive bonuses; some of the educators received prison sentences (Fausset and Blinder 2015). In Biscayne Park, Florida, a police department systematically made false arrests to clear outstanding cases and improve crime statistics (Caron 2018).

All three cases show that cheating and data manipulation are real threats to performance systems. While the severity and scope of the system dysfunctions in the three examples are extreme, rewards for improvements in many settings are likely to also incentivize unintended side-effects, although often more modest and undetected. The literature supports such thinking, as it provides evidence for the existence of performance gaming and cheating across policy areas and countries. Scholars also agree that performance gaming is a systemic issue with detrimental consequences for clients and citizens, hence warranting further research that contributes to understanding this phenomenon better (Benaine and Kroll 2020; Bohte and Meier 2000; Eterno and Silverman 2012; Hood 2006; Jacob and Levitt 2003; Koning and Heinrich 2013; Li 2015; Radin 2006).

Although previous research has focused on configurations of the performance system itself as an important factor (Eterno and Silverman 2012; Heinrich and Marschke 2010; Radin 2006), little is understood with regard to the heterogeneity of behaviors under the same performance regime. In other words, once the performance system variable is held constant, why are some employees more likely to engage in dysfunctional behaviors than others? Furthermore, a great deal of research examined gaming behaviors that can be observed or reconstructed using case studies and post-hoc data analysis, but the least is known about the most radical type—outright cheating (falsifying or fabricating data).

Our study addresses these research gaps. We identify three factors that may explain employees' performance cheating: prosocial impact, job stress, and red tape. First, when having to choose between serving the performance system as opposed to the clients, a high level of perceived prosocial impact may make employees choose the latter over the former. Second, we argue that job stress depletes employees' cognitive capacity to resist unethical choices, hence fostering cheating. Third, working around red tape provides a self-serving justification to employees why unethical behavior may be tolerable, which increases performance cheating.

In order to test these hypotheses, we screen through a sample of almost 10,000 potential subjects to identify about 1,000 individuals who work in public administration and use performance data as a part of their job. As such, we conduct a list experiment (Imai 2011) to estimate the prevalence of performance cheating. In a nutshell, we randomly divide subjects into two groups and ask individuals about the number of performance practices they have previously engaged in (e.g., "explain performance data, so my supervisor understands them"). One practice is "altering data, so that the results look better." This is our sensitive item, capturing performance cheating, which was only included for one of the two groups. By knowing the number of items individuals in both groups (one with and the other without the sensitive item) agreed with, we can estimate the agreement with the cheating item, a number that serves as our study's dependent variable (Glynn 2013; Imai 2011). We measure all other variables using a survey instrument and employ item-count technique regression analyses (Blair and Imai 2012) to determine each variable's impact on performance cheating. We find that all three factors reinforce performance cheating among public employees.

Manipulating Effort and Data in Public Organizations

In line with previous work, we define gaming as generating positive performance data without achieving the actual objective behind the indicator (Benaine and Kroll 2020; Heckman et al. 2011), and distinguish between two types: the manipulation of effort and the manipulation of data. The former refers to changes in behaviors in order to comply with the performance system, while the latter is about alterations in data without actual behavioral change. However, in our categorization of gaming, we introduce a second dimension—the severity of the manipulation regarding its detriment for clients and citizens—and, hence create a two-by-two typology of performance gaming (see figure 1). We use this typology to help structure the literature on the subject and define more clearly the specific gaming type our study focuses on, which is performance cheating.

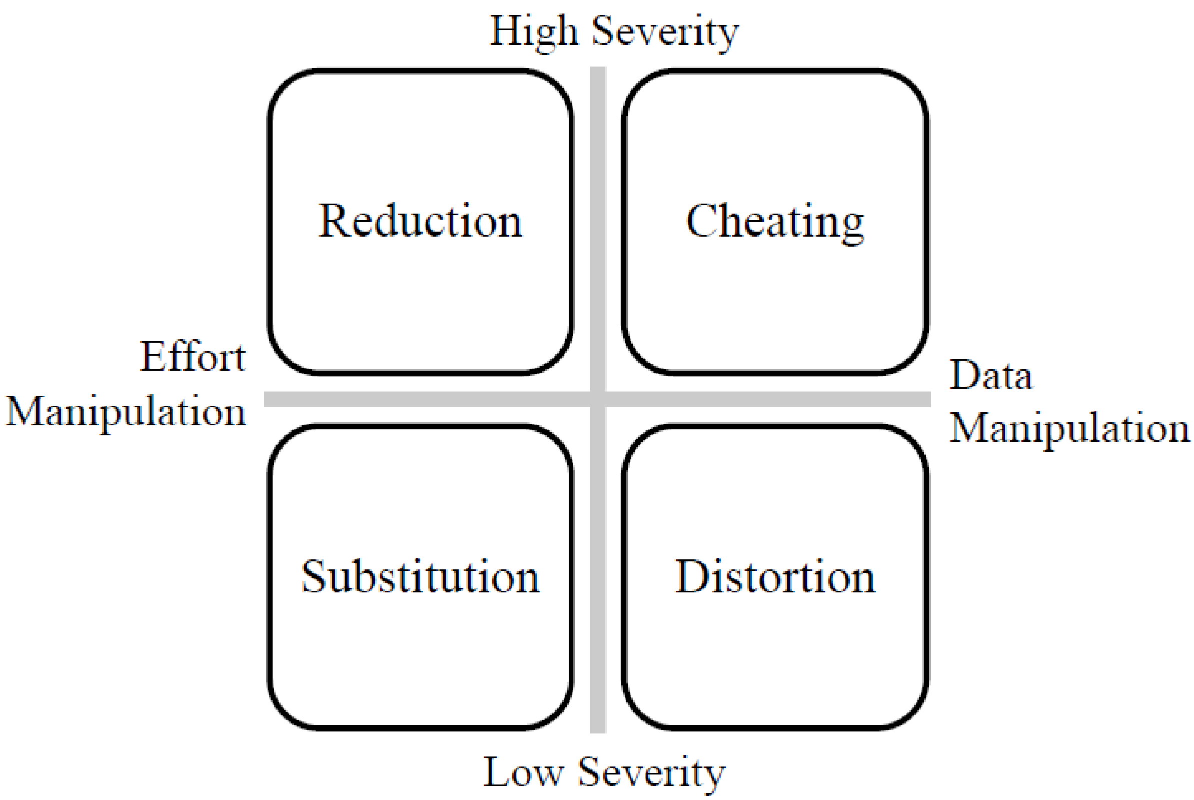


Figure 1: Performance Gaming Typology

A low-severity type of effort manipulation is substitution, which has also been referred to as goal displacement, cherry-picking, or cream-skimming (Radin 2006). Here, organizations or individuals shift resources and attention away from unrewarded tasks to rewarded ones. We consider the severity of this effort as relatively low because effort substitution could still yield some performance improvements, however, only in the areas selected. Koning and Heinrich (2013) examine private social welfare providers in the Netherlands and find that they better serve more easily employable workers and focus on job placements rather than duration. Benaine and Kroll (2020) show that high schools in the U.S. trade off unrewarded for rewarded indicators of learning, particularly when serving disadvantaged students. Gerrish (2017), in his study of the Child Support Performance and Incentive Act, documents a positive effect on one rewarded and a negative effect on one unrewarded outcome. Kelman and Friedman (2009) examine hospital trusts in England but find no systematic evidence for effort substitution when analyzing performance against a predefined emergency room wait-time target.

A high-severity type of effort manipulation is reduction. Here, organizations or individuals are likely to generally reduce, or hold back, effort because targets were set too high (or not high enough) and are thus not motivating, or improvements will lead to higher future targets. Behn (2012) points out that most agencies, whose performance is compared with similar entities, may slack off if the target is set so high that only the best performers can achieve it. Bevan and Hood (2006) study the English healthcare system and find evidence for the crowding of performance towards the target. While more ambulance trusts had met the eight-minute response target compared to a decade ago when it did not receive as much attention, very few actually exceeded this threshold. Another example of changing service provision in a way that reduces quality in order to meet performance requirements includes allegations that patients waited in ambulances rather than waiting rooms to meet the formal wait-time target (Bevan and Hood 2006).

A low-severity type of data manipulation is distortion, including the selection of favorable indicators or the highly distorted interpretation of numbers introducing a great deal of bias. Courty and Marschke (2007) provide the example of a federal job training program in the U.S., where training centers kept trainees on the books until they had actually found a job, or they changed their status in the time period in which they could do the least harm to their overall metrics. Bohte and Meier (2000) document a distortion strategy they call "biased sampling," where schools liberally extend definitions of students with special needs to exempt low-performing students from standardized test score calculations. Bevan and Hood (2006) provide more examples of data distortion from the English health care system, including the reclassification of response time metrics in order to show target compliance.

A high-severity type of data manipulation is cheating. Such behavior includes the fabrication and falsification of data that goes beyond "creative" use and is often illegal. The examples provided at the beginning of the article fall in this category: manipulating waitlists or falsifying students' test scores (Caron 2018; Fausset and Blinder 2015). Our study is on this type of performance gaming and, hence, different from most of the research referenced in the paragraphs above. While previous work considers cheating to be a substantive threat to performance systems and their ability to create public value, most previous scholarship has been concerned with the manipulation of effort or data distortion (Eterno and Silverman 2012; Hood 2006; Jacob and Levitt 2003). Rather than drawing on case studies or data analysis techniques to reconstruct gaming behavior, our work will use an experimental approach and a large-N sample that cuts across organizations. In what follows, we theorize about potential determinants of performance cheating within government accountability systems.

Determinants of Performance Cheating

Research on unethical intentions and behaviors is growing in the area of public administration (e.g., Bellé and Cantarelli 2017a; Christensen and Wright 2018; Olsen et al. 2019). Since corrupt and dishonest behaviors are present but difficult to observe in public management practice, the most recent studies use experimental methods to measure such behavior and establish causal relationships. What these studies also have in common is the notion that much can be learned from other disciplines' theories and methods related to dishonest behavior. In this vain, Bellé and Cantarelli (2017b) conducted a meta-analyses, whose main objective is to systemically review evidence for causes of unethical behavior published in journals outside of our discipline and make recommendations of future research avenues in public administration research.

In order to propose a testable model of performance cheating, we build on the emerging scholarship on unethical behavior in public administration as well as the longstanding research in other disciplines, most notably behavioral economics, social and cognitive psychology, management, and organizational behavior (e.g., Ayal et al. 2015; Kish-Gephart, Harrison and Treviño 2010; Zahra, Priem and Rasheed 2007). Unlike in previous work, our theoretical interest is not in unethical behavior in general but explicitly cheating behavior within public performance systems, that is, the altering of performance information in government accountability regimes (Benaine and Kroll 2020; Heinrich and Marschke 2010). Consequently, we draw on research on unethical and dishonest behaviors more broadly and adapt previous findings to the case of public performance systems specifically.

The first factor in our model is prosocial impact, which we understand as the opportunity to benefit the lives of clients or customers through one's job (Bellé 2014; Grant 2012). We find only a small amount of empirical research showing that prosocial impact may curb unethical behavior. Scattered evidence includes experimental research demonstrating that when the negative impact on others becomes more visible, people engage less in unethical behavior (Yam and Reynolds 2016). Similarly, individuals are more likely to show behaviors consistent with being a “good citizen” if inspirational messages are delivered to them by beneficiaries of the individuals’ work, particularly when highlighting prosocial impact (Grant and Hofmann 2011).

In contrast, there seems to be more empirical support for the notion that prosocial impact may reinforce unethical behavior rather than prevent it. In their review, Bolino and Grant (2016:618) argue that prosocial values can yield unethical behavior, particularly if such behavior benefits colleagues or clients. Christensen and Wright (2018) run a set of experiments with students but find no evidence for any sort of link between primed prosocial values and unethical behavior. Bellé and Cantarelli (2017a) examine students as well as public employees in two experiments and find consistent support for a reinforcing effect of prosocial impact on unethical behavior. One explanation of their findings is that individuals whose performance is related to prosocial consequences are more likely to engage in misconduct since what they do benefits other people (Erat and Gneezy 2012).

The ability to affect others’ lives positively can trigger cheating if individuals have to trade-off compassion against neutrality or lawfulness. Focusing strictly on client wellbeing could come at the expense of drifting away from the organization’s mission or policies and engage in illegal rule-bending (Schott and Ritz 2018). With respect to performance systems, this could mean that if aware of one’s prosocial impact, public employees may falsify impact data if that helps to direct organizational attention and resources to the people they serve. Such behavior includes engaging in moderate levels of gaming to create a performance “buffer” that could be used later to accommodate the most difficult-to-serve clients (Heckman et al. 2011).

Theoretically, such data fabrication could include over- or under-reporting of performance, depending on whether additional resources are given to the best performing units or clients in need.

H_1: Prosocial impact is associated with more performance cheating.

Scholarship on performance systems has made the case that a negative side-effect of using incentives is dysfunctional responses, such as gaming and cheating, and that these are more likely to occur, the more significant the consequences attached to the systems (e.g., Bevan and Hood 2006; Bohte and Meier 2000; Heinrich and Marschke 2010). As such, a great deal of performance pressure can motivate data manipulation. Our argument takes this point one step further, in that we propose that job stress generally—and not just pressures created through an accountability system—will encourage performance cheating among public employees.

Here, job stress refers to the anxiety experienced at work. In line with Brooks and Schweitzer (2011:44), we understand anxiety as a “state of distress and/or physiological arousal in reaction to stimuli, including novel situations and the potential for undesirable outcomes.” Job stress can have many causes, including powerlessness, conflict, poor human relations, punishment, isolation, and uncertainty (Bazerman and Tenbrunsel 2011). We see two central mechanisms that explain why job stress can lead to unethical behavior (Gino et al. 2011; Kouchaki and Desai 2015; Selart and Johansen 2011). The first one is evolutionary and falls into the category of “fight or flight.” When put in a stressful situation, the human mindset prioritizes survival over morality, and hence individuals will do what it takes to come out on top, including cheating. The second mechanism is more cognitive in nature. Moral decisions require self-regulation, which, in turn, is an extra effort that necessitates cognitive capacity. However, stress depletes such capacity, making more simplistic, self-interested choices likely to occur.

Along these lines, previous research shows that stress resulting from being under time pressure can lead to unethical behavior (Bellé and Cantarelli 2017b). In multiple experiments, Kouchaki and Desai (2015) find that anxiety makes people feel threatened, which then makes it more likely for them to engage in unethical behavior. Similarly, Selart and Johansen (2011) report based on two studies that leadership stress curbs ethical acting among a sample of decision-makers. Overall, we expect that job stress will increase the likelihood that employees behave unethically. If public employees work under accountability regimes, then behaviors including collecting, reporting, and analyzing performance data offer opportunities for them to cheat.

H_2: Job stress is associated with more performance cheating.

When theorizing about rules and cheating, a significant stream within the literature takes an institutional perspective, in that formalized processes within the bureaucracy (e.g., monitoring, audits, and "sunshine" regulations) can prevent unethical conduct (for extensions of this point and evidence, see Olsen et al. 2019). Therefore, Kish-Gephart et al. (2010), in a meta-analysis, show that a principle-based ethical climate—where employees tend to follow organizational rules rather than deviate from formal norms—is negatively related to unethical intentions and behaviors.

We take a different perspective on formalized institutions and specifically turn to the negative effects of ineffective rules. We suggest that if rules are perceived as counterproductive, employees will apply more personal discretion and work around formal systems. If such behavior is learned within an organization and has become the norm and widespread practice, then it is likely that it will also determine how seriously employees treat performance data and systems. In order to develop this argument, we draw on the literature on red tape, which is about “rules that remain in force and entail a compliance burden for the organization but make no contribution to achieving the rules’ functional objective” (Bozeman and Feeney 2011:44). Red tape is known to yield several negative consequences in government agencies, including impeding change and performance as well as frustrating employees (Brewer and Walker 2010; DeHart-Davis and Pandey 2005).

In light of the above, how rules are perceived by public employees determines the extent to which they are followed. Rules perceived as effective (“green tape”) are more likely to create abidance among employees, while the opposite suggests that perceptions of little effectiveness (“red tape”) can lead to rule-bending (DeHart-Davis 2009). Individuals tend to maintain a positive self-image and, thus, do not engage in unethical behavior (e.g., deviating from rules and standards) as the default option. Rather, this is the exception and requires a critical level of self-serving justification (Ayal et al. 2015). Shalvi et al. (2015) argue that when making unethical choices, individuals deal with a type of ethical dissonance, where they want to benefit from wrongdoing but at the same time still feel good about themselves. In that sense, self-serving justifications reduce the psychological costs of unethical behavior.

Labeling rules as red tape justifies ignoring them, and informal norms largely determine how employees perceive and act on formal rules (Borry 2017; Bozeman and Feeney 2011). Public employees who perceive a great deal of “bad rules” in their organization may also be more likely to use such a stigma as an explanation for why they work around formal systems. This may include rules and regulations generally, but also performance practices specifically. Research on different types of rule violations, such as rule-bending and rule-breaking (DeHart-Davis 2017), broadly mirrors the differential intensities of dysfunctional data use we previously alluded to—distortion and cheating. Performance systems embedded in an environment of dysfunctional rules are less likely to be useful and more likely to be ignored, treated as a compliance exercise, or even manipulated. After all, learned behavior such as doing what it takes to deal with red tape may expand into other domains dominated by formal standards and requirements established through performance systems.

H_3: Red tape is associated with more performance cheating.

Research Design

The List Experiment

Manipulating performance data is not a behavior like any other. It is inconsistent with organizational policies, can be perceived as unethical, and in some cases is even a punishable offense. Hence, designing a survey that directly asks public employees whether they engage in cheating is not a promising research design. Since participants have strong incentives to refuse to answer such a survey honestly, we would have to assume that the results do not give a truthful estimation of the prevalence of cheating, and the subsequent analysis of influence factors would be strongly biased.

In order to overcome this limitation of survey research, methodologists, especially in political science, have developed several methods to design surveys that ensure participants' confidentiality and increase the accuracy of the responses, and therefore allow to estimate the prevalence of sensitive behaviors or attitudes accurately (Nuno and St. John 2015). We decided to implement a so-called list experiment, also known as the item-count technique (Imai 2011). It is easy for participants to understand its application, and recent methodological developments allow the analysis of influence factors on the affirmation to the sensitive question (Blair and Imai

2012; Imai 2011). The method has been used in a variety of contexts to address various sensitive behaviors and attitudes such as racial prejudice, sexual risk behaviors, and citizens' support of body cameras worn by police officers (e.g., Blair and Imai 2012; Bromberg, Charbonneau and Smith 2018; Hubbard, Casper and Lessler 1989).

The list experiment ensures the confidentiality of the respondents by asking them to indicate how many statements from a list of statements they agree with, rather than scoring their agreement with individual items. The list contains one statement about the sensitive attitude or behavior of interest and some other non-sensitive statements. As long as respondents do not agree with all of the statements, it remains confidential, which statements they agreed with. The sample is randomized into two groups to draw inferences from the resulting data. One group is asked to indicate the number of statements they agree with on the full list, while the second group gets a list without the sensitive item. The difference in the mean number of agreed items between the group that saw the full list and the control group that saw the list without the sensitive statement can, therefore, be treated as the prevalence of the sensitive attitude or behavior (Glynn 2013).

In order to estimate a truthful prevalence of performance cheating in the public sector, participants were asked to indicate how many of the following four control items they agree with:

As a part of my job, I …

(1) Make decisions and allocate funds based on performance data
(2) Collect and consolidate performance data from colleagues or stakeholders
(3) Explain performance data so my supervisor understands them
(4) Correct errors and clean data but don't really make any decisions

Participants randomly assigned to the treatment group saw a list with the following additional item that reflects performance cheating:

(5) Alter data so that the results look better

Selecting the control items, we followed the recommendations given in the literature (e.g., Blair and Imai 2012; Glynn 2013; Hubbard et al. 1989) and—among other considerations—selected two negatively correlated items (items 1 and 4) to avoid participants agreeing with all statements. The sensitive item captures performance cheating, a high-severity type of data manipulation. Altering data includes the fabrication and falsification of performance information and goes beyond "creative" data use.

To test the hypothesized effects of prosocial impact, job-related anxiety, and red tape on performance cheating, we employ techniques developed by Imai (2011), which enable researchers to go beyond just estimating the prevalence of a sensitive attitude or behavior and allow the use of multivariate regressions to test the effect of independent variables on the agreement with the sensitive item. Among the available estimators, we use maximum likelihood estimation as it is the most efficient estimator providing the most accurate coefficients and standard errors (Blair and Imai 2012). In such a model, the dependent variable is the likelihood that a respondent agrees with the sensitive item. Since this information is unknown, the likelihood to agree with the sensitive item is modeled from the joint probability of agreeing with the sensitive item and the number of control items (Tsai 2019).

Data

Since it is challenging to recruit participants for sensitive research topics, we decided to use a convenience sample of participants recruited within Amazon Mechanical Turk (MTurk) (Stritch, Pedersen and Taggart 2017). It is noteworthy that our theoretical interest is in performance cheating behavior that can occur at the employee or clerk level, not necessarily the level of the decision-maker. Empirical research shows that deviations of the representativeness of the MTurk population from the general population are limited (Huff and Tingley 2015) and that experiments conducted on MTurk in general lead to the same results as population-based random samples (Coppock 2019).

To make sure that respondents were public employees and work with performance data, we applied a two-step procedure. First, we submitted our survey to MTurk as a screening survey offering $0.10 for participation and $1.75 for an additional survey if participants were eligible based on their answers in the screening survey. The screening survey asked participants to indicate what proportion of their working time they spend on five different tasks, one of which was “dealing with data and performance information.” In addition, participants had to state if and in which sector they were employed (see Appendix A). Participants had to indicate that they worked in the public sector and spent at least five percent of their time working with data to be invited to the main survey. We allowed participants to take the survey until we reached the recommended sample size for list experiments of about 1,000 participants.

Using this procedure, we recruited 9,909 participants who filled out the screening survey, out of whom 1,005 met the inclusion criteria and participated in the main survey. We excluded 14 participants from the analysis as their open text answers to a question about their job title indicated that they did not pay attention to the survey or were not eligible in the first place. In addition, 27 participants did not answer the two attention checks correctly and were therefore excluded from the analysis. This final sample stands at 964 participants. Participants work in all parts of the public sector, with 36.3 % in education; 21.5 % in government including local government; 14.7 % in health and social policy; 10.6 % in defense, criminal justice, and national security. The remaining 16.9 % are distributed in ten additional areas including, among others, community and society, environment, and transportation. Most respondents' job titles fall in the following categories: educators, administrators, clerical staff, auxiliary services, law enforcement officers, nurses and nursing assistants, information technology personnel, and social workers. While our sample is large relative to other studies in our area, we encourage replications with even larger data sets because list experiments tend to be fairly sensitive to changes in sample size.

Operationalizations

Besides capturing the dependent variable, performance gaming, using the list experiment, we measured the independent variables to test the proposed hypotheses and some control variables in the standard survey procedure. The full set of measures is listed in Appendix B. We used four items ($\alpha = .93$) by Grant and Campbell (2007) to assess participants' perception of the prosocial impact of their job (Moynihan, Pandey and Wright 2012). The five job-related anxiety items (Cronbach's $\alpha = .83$) were developed by Parker and DeCotiis (1983) and like subsequent variables measured on a scale seven-point Likert scale (1 = not at all; 7 = very much). Red tape was operationalized as a single item derived from Rainey, Pandey and Bozeman (1995), previously used by, among others, Pandey and Rainey (2007).

While most of our variables of interest are at the (behavioral) micro level, we do control for what we think is the most important macro-level confounding factor: the organization's focus on performance improvement. Specifically, we use a four-item scale ($\alpha = .77$) that captures the consequentiality of the performance system: the role of performance data in meetings, appraisals, and reward structures. A second central variable is the proportion of their working time participants spend “dealing with data and performance information.” All multi-item measures were aggregated as mean indices. Finally, we controlled for participants’ age, gender, and education. It is noteworthy that item-count regressions are prone to issues related to statistical power, which is why the number of control variables we can include in our model is limited.[1]

Results

Table 1 shows the descriptives for the independent and control variables. The first issue we are interested in is how common performance cheating is. The list experiment method provides a straightforward answer to that question. Since participants are randomly assigned to the treatment or control group and the only difference between the groups is the number of items in the list (four vs. five), any difference between the groups can be attributed to the sensitive item. By subtracting the mean number of agreed items in the treatment group from the mean of the control group (difference-in-means estimation), we receive an unbiased estimation of agreement with the sensitive item (Blair and Imai 2012). In our sample, participants in the treatment group on average agreed with 2.0062 items and participants in the control group on average with 1.9459 items. The estimated prevalence of performance cheating is 0.0603 or, considering that a

[1] At the end of the survey, we also employed a question that tried to measure performance cheating directly. However, due to a host of issues related to this measure’s validity (including monotonicity, Aronow et al. 2015), we refrain from using it for any additional analyses. Additional tests suggest that subjects may have perceived the direct measure as an attention check of their previous response to the item list (our indirect measure), which is commonly used in the Amazon Mechanical Turk environment, but in our case invalidates the responses to the direct instrument.

full one-item difference would constitute a 100 % agreement with the cheating item, 6.03 %.

Table 1: Descriptive statistics

Variable	**n**	**Mean**	**St. Dev.**	**Min**	**Max**
Prosocial impact	964	5.488	1.211	1	7
Job-related anxiety	964	3.595	1.510	1	7
Red tape	964	5.892	2.500	0	10
Organizational performance focus	964	4.137	1.069	1	6
Performance data exposure	964	22.535	17.796	5	100
Age	964	35.754	9.982	19	74
Gender (1 = female)	964	0.460	0.499	0	1
Education	964	3.930	1.092	1	5

Table 2 displays the results of a multivariate regression on the likelihood of agreeing with the sensitive item using an unconstrained[2] maximum likelihood estimator.[3] The model finds support for hypotheses 1. Prosocial impact encourages performance cheating, although the effect is only significant at the 10% level (b = 1.070, p = .060). We also find support for a reinforcing effect of more job-related anxiety on cheating (b = 0.437, p = .022). Additionally, participants reporting a high level of red tape in their organization are also more likely to fudge performance data (b = 0.313, p = .021).

[2] Following Imai's (2011:411) advise, we estimated an unconstrained and a constrained model and compared them using a likelihood ratio test. As the test was significant, we use the unconstrained model.

[3] The full model including the estimated influence of the independent variables on the response to the control items can be found in Appendix C.

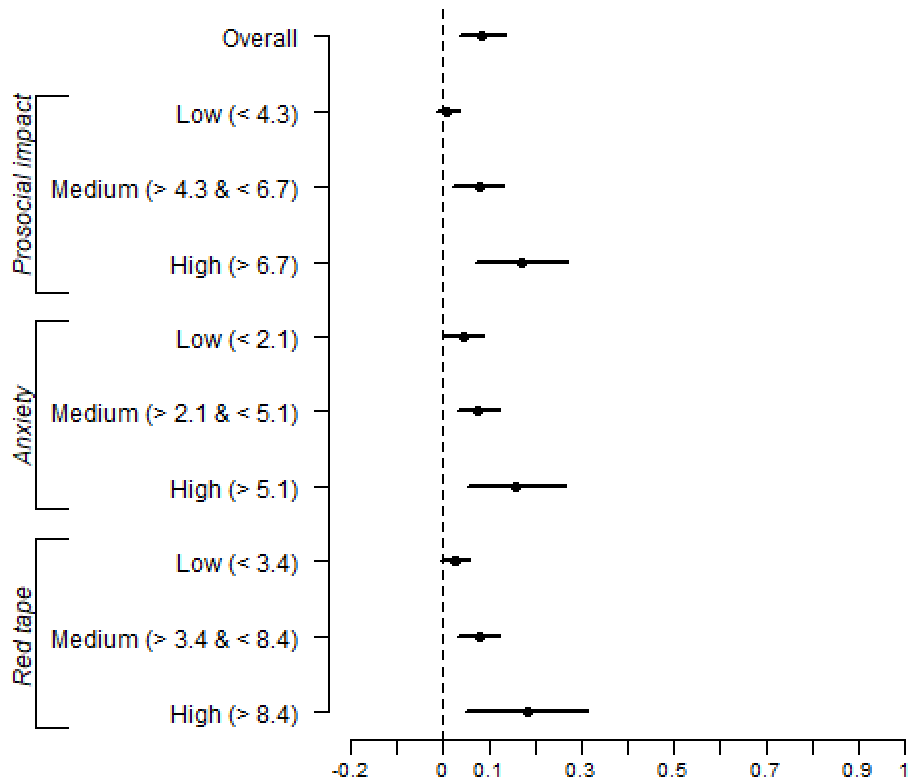
Table 2: Multivariate analysis of list experiment

Variable	**Est.**	**p**
Prosocial impact	1.070	0.060†
Job-related anxiety	0.437	0.022*
Red tape	0.313	0.021*
Controls		
Organizational performance focus	0.229	0.590
Performance data exposure	0.003	0.886
Age	0.089	0.017*
Gender (1 = female)	1.246	0.077†
Education	-0.740	0.003**
Intercept	-14.964	0.000***
n	964	

Note: Estimated coefficients from the item count technique regression models, where the sensitive item is whether or not participants 'alter data so that the results look better' in their job. Coefficients are based on logistic regression models with unconstrained maximum likelihood estimation. † $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

Among the control variables, there are two significant effects. Older participants are more likely to cheat (b = 0.089, p = .017), just as more educated participants are (b = -0.740, p = .003). The effect of gender on performance cheating indicates that women are more likely to cheat (b = 1.246, p = .077), but this effect is only significant on a 10 % significance level and contradicts previous findings (Portillo and DeHart-Davis 2009). The performance focus of the organization, and the exposure to the performance data, does not affect the likelihood of cheating once all factors have been accounted for. In Appendix D, we examine the common assumptions that list experiments are based on but find no violations.

Figure 2: Estimated proportions of participants cheating performance data by values of the independent variables.



Dots represent estimated proportions of respondents cheating performance data, and lines are 95 % confidence intervals from the binomial logistic regression model in Table 2. The graph emphasizes that performance cheating is higher for participants with higher values of perceived prosocial impact, job-related anxiety, and red tape.

For a more intuitive understanding of the substance of the effects, we inserted some illustrative numbers for the independent variables and created predictive values from our model in Table 2. We used one standard deviation below and above the mean as the cut-off to illustrate the proportion of the sample estimated to agree with the sensitive item. Figure 2 displays the resulting predictions together with their 95 % confidence intervals. The figure shows that agreement with the sensitive item is substantial for employees who scored high on our three independent variables. Of those who report a high perceived prosocial impact, 16.78 % also manipulate performance data, as opposed to only 1.00 % of those with low prosocial impact. Among those with high job-related anxiety, 15.87 % also engage in cheating (versus 4.23 % of those with low job-related anxiety). Moreover, of those who report high levels of red tape in their organization, 18.06 % alter performance data as well (versus 2.73 % of those with low levels of red tape).

Discussion

Our findings suggest that with respect to government performance systems, employees who are aware of their jobs' prosocial impact are also more likely to manipulate performance data. Our explanation is that serving clients can be at odds with satisfying formal performance requirements, which is why workers may find themselves in situations in which they have to trade off both objectives (Bellé and Cantarelli 2017a; Bolino and Grant 2016). As expected, job stress and red tape increase performance cheating. While the negative effects of both variables have been prominently featured (Bellé and Cantarelli 2017b; Bozeman and Feeney 2011; Brewer and Walker 2010; Kouchaki and Desai 2015), we extend such observations to the specific case of performance systems. As such, stress may put employees in a position in which reporting "good news" is often rendered more important than doing the "right thing." Red tape, in contrast, can provide a self-serving justification as to why cheating should be considered acceptable.

As job stress has been featured the least in the public management literature, we elaborate here a little more on this factor to facilitate the understanding of our findings as well as the design of future studies. Most broadly, job stress is likely the result from overload, conflict, or ambiguity in the workplace (Dewe, O'Driscoll and Cooper 2010). A fair amount of research has identified the misfit between person and environment as a crucial antecedent of job stress.

Specifically, attention has been given to the fit between preferred and received stimuli (e.g., in terms of needs and actual supply, see Yang, Che and Spector 2008) as well as abilities and demands. Other research has been concerned with contingency factors that may moderate the detrimental effect of straining job demands, pointing to social support and one's control over demands as promising moderators (Daniels et al. 2008). While job stress is a predictor in our model, it may also serve as an intervening variable that links broader job-design choices to cheating behavior.

We see two specific strengths of our study: (1) we employ an experimental method to capture cheating that is reliable and superior to subjective self-reports of sensitive behaviors or other more direct observational methods; (2) our experimental measure of cheating is based on reports of employees' behavior in the real workplace, rather than hypothetical assignments in a more artificial laboratory environment. While the latter approach has the advantage that researchers have full control over the experiment, its context, and conditions (Olsen et al. 2019; Shalvi et al. 2015), lab settings also tend to lack some degree of external validity.

We acknowledge the limitation that the sample we work with is a group of lower-level employees and clerks from a cross-section of public organizations, about whose administrative features we know relatively little. However, similar to previous survey research in this area (e.g., Kroll and Vogel 2014; Moynihan and Pandey 2010), we do control for differences in organizations' performance systems, which are highly relevant in answering our research question. Furthermore, recent scholarship has shown that MTurk samples tend to be fairly representative and of utility to behavioral researchers (Coppock 2019; Huff and Tingley 2015; Stritch et al. 2017). In addition, we find it noteworthy that our sampling choice matches our theoretical interest. In fact, while most traditional research on performance cheating and gaming conceptualizes the organization's performance system at the level of the decision-maker ("dysfunctional performance information use"), we see a gap in this literature, as it tends to disregard behaviors below the level of use, including the collection, reporting, and analysis of performance data. Other than most previous work based on samples from one or a small number of public organizations, using MTurk, we were able to cast our net wide to recruit public employees all across North America.

Along these lines, we further believe that our sampling strategy was not prone to severe selection effects. Due to a screening survey that included four employment options (being unemployed; working for a private company or organization; a nonprofit organization; or public administration/government) and the task of ranking several work behaviors—only one of which referred to the handling of performance data—potential participants had virtually no way of knowing how to respond to the screening survey if it was their intention to be included in the main survey. Recent research shows that such a procedure avoids biased inference that occurs when the eligibility criteria are known to potential participants (Siegel and Navarro 2019).

Common source bias is not a major concern for our study. The question scaling for dependent and independent variables was entirely different, and social desirability bias is minimized due to two of the main features of list experiments (Jakobsen and Jensen 2015): subjects a) are not aware of the sensitive item and b) even if they are, believe that the researchers will not be able to determine how they rated the sensitive item (Blair and Imai 2012; Glynn 2013). Essentially, we asked respondents how many items they agree with, not which ones. Furthermore, the dependent variable is not directly based on subjects' self-reports; rather, individual cheating scores are estimated based on overall probabilities and distributions. Although our design has a higher external validity than many other studies researching gaming in an artificial laboratory context, this advantage also has a downside. A limitation is that the relationship between the independent variables and gaming is correlational and therefore needs additional research to prove causality.

Conclusion

In this article, we use a list experiment to measure employees' performance gaming—an instrument that addresses the limitations of direct self-reports of sensitive behaviors via traditional surveys or interviews. Our main findings that prosocial impact, job stress, and red tape tend to make performance cheating among employees more likely suggests the following for management practice. First, in order to avoid that employees have to trade-off making a prosocial impact against meeting organizational targets, performance systems need to be aligned with the agency's prosocial values. One way to do this is to ensure that key performance indicators capture all aspects of the organization's mission, even including non-mission based measures that capture dimensions such as social equity, procedural justice, or social trust. Second, in addition to the many known harmful effects of job stress, we find that anxiety at work increases the odds of performance cheating. To mitigate this effect, agencies may want to promote practices related to a work-life balance and consider employees' shortcomings as opportunities for growth and learning. Third, since red tape reinforces cheating, organizations need to audit existing rules (specifically those pertaining to the performance system) to weed out red tape but also engage in a dialogue with employees about the utility and effectiveness of existing rules to manage rule perceptions.

Our findings also contribute to the maturing of gaming theory. As research areas mature, theories become more sophisticated and causal chains tend to be extended. This includes moving from studying narrow relationships between variables that are conceptually closely tied to broader, more generic social relationships, thereby improving our theories' falsifiability and utility and moving from prediction to explanations (Bacharach 1989). Research on the purposeful use of performance information has already gone through such a process of maturation, following Moynihan and Pandey's call, who stated in 2010 (p. 862): The "narrow construction of independent variables helps to answer the question: 'how can organizations foster performance information use?' But we know less about how more commonly occurring and broadly constructed organizational concepts matter to performance information use." Yet, theories of performance gaming in government organizations are still largely focused on variables related to the incentive system itself (with a few exceptions) (Bohte and Meier 2000; Eterno and Silverman 2012; Heinrich and Marschke 2010; Radin 2006).

With our study, we want to contribute to the extension of existing causal chains in the performance gaming literature. We turned to variables that have been found to provide meaningful explanations of public administration behavior (e.g., prosocial impact and varying perceptions of rules; Bellé and Cantarelli 2017b; DeHart-Davis 2009) or were featured in theories outside of our discipline (e.g., job-related anxiety; Brooks and Schweitzer 2011), but have received little consideration in work on government performance systems. Understanding such broader mechanisms behind performance gaming can offer additional opportunities for behavioral interventions (e.g., Ayal et al. 2015; Shalvi et al. 2015) to nudge employees in directions that are less morally controversial and potentially harmful to the public interest.

References

Aronow, Peter M., Alexander Coppock, Forrest W. Crawford, and Donald P. Green. 2015. "Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology* 3(1):43–66. doi:10.1093/jssam/smu023.

Ayal, Shahar, Francesca Gino, Rachel Barkan, and Dan Ariely. 2015. "Three Principles to REVISE People's Unethical Behavior." *Perspectives on Psychological Science* 10(6):738–41. doi:10.1177/1745691615598512.

Bacharach, Samuel B. 1989. "Organizational Theories: Some Criteria for Evaluation." *The Academy of Management Review* 14(4):496. doi:10.2307/258555.

Bazerman, M. H., and A. E. Tenbrunsel. 2011. *Blind spots: Why we fail to do what's right and what to do about it*. Princeton, NJ: Princeton University Press.

Behn, Robert D. 2012. "Motivating and Steering with Comparative Data." *International Public Management Review* 13(1):21–37.

Bellé, Nicola. 2014. "Leading to Make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation." *Journal of Public Administration Research and Theory* 24(1):109–36. doi:10.1093/jopart/mut033.

Bellé, Nicola, and Paola Cantarelli. 2017a. "Do Ethical Leadership, Visibility, External Regulation, and Prosocial Impact Affect Unethical Behavior?: Evidence From a Laboratory and a Field Experiment." *Review of Public Personnel Administration*. doi:10.1177/0734371X17721301.

Bellé, Nicola, and Paola Cantarelli. 2017b. "What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research." *Public Administration Review* 77(3):327–39. doi:10.1111/puar.12714.

Benaine, S. L., and Alexander Kroll. 2020. "Explaining Effort Substitution in Performance Systems: The Role of Task Demands and Mission Orientation." *Public Management Review* 22(6):813–35. doi:10.1080/14719037.2019.1604794.

Bevan, Gwyn, and Christopher Hood. 2006. "What's measured is what matters: Targets and gaming in the English public health care system." *Public Administration* 84(3):517–38. doi:10.1111/j.1467-9299.2006.00600.x.

Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(01):47–77. doi:10.1093/pan/mpr048.

Bohte, John, and Kenneth J. Meier. 2000. "Goal Displacement: Assessing the Motivation for Organizational Cheating." *Public Administration Review* 60(2):173–82. doi:10.1111/0033-3352.00075.

Bolino, Mark C., and Adam M. Grant. 2016. "The Bright Side of Being Prosocial at Work, and the Dark Side, Too: A Review and Agenda for Research on Other-Oriented Motives, Behavior, and Impact in Organizations." *The Academy of Management Annals* 10(1):599–670. doi:10.1080/19416520.2016.1153260.

Borry, Erin L. 2017. "Ethical climate and rule bending: How organizational norms contribute to unintended rule consequences." *Public Administration* 95(1):78–96. doi:10.1111/padm.12304.

Bozeman, Barry, and Mary K. Feeney. 2011. *Rules and Red Tape: A Prism for Public Administration Theory and Research*. Armonk, New York: M.E. Sharpe, Inc.

Brewer, Gene A., and Richard M. Walker. 2010. “The Impact of Red Tape on Governmental Performance: An Empirical Analysis.” *Journal of Public Administration Research and Theory* 20(1):233–57. doi:10.1093/jopart/mun040.
Bromberg, Daniel E., Étienne Charbonneau, and Andrew Smith. 2018. “Body-Worn Cameras and Policing: A List Experiment of Citizen Overt and True Support.” *Public Administration Review* 78(6):883–91. doi:10.1111/puar.12924.
Brooks, Alison W., and Maurice E. Schweitzer. 2011. “Can Nervous Nelly negotiate? How anxiety causes negotiators to make low first offers, exit early, and earn less profit.” *Organizational Behavior and Human Decision Processes* 115(1):43–54. doi:10.1016/j.obhdp.2011.01.008.
Caron, Christina. 2018. “Three Florida Police Officers Are Sent to Prison for False Arrests.” (https://www.nytimes.com/2018/10/19/us/florida-police-officer-false-arrests.html).
Christensen, Robert K., and Bradley E. Wright. 2018. “Public service motivation and ethical behavior: Evidence from three experiments.” *Journal of Behavioral Public Administration* 1(1):1–8. doi:10.30636/jbpa.11.18.
Coppock, Alexander. 2019. “Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach.” *Political Science Research and Methods* 7(3):613–28. doi:10.1017/psrm.2018.10.
Courty, Pascal, and Gerald Marschke. 2007. “Making Government Accountable: Lessons from a Federal Job Training Program.” *Public Administration Review* 67(5):904–16. doi:10.1111/j.1540-6210.2007.00777.x.
Daniels, Kevin, Nick Beesley, Alistair Cheyne, and Varuni Wimalasiri. 2008. “Coping processes linking the demands-control-support model, affect and risky decisions at work.” *Human Relations* 61(6):845–74. doi:10.1177/0018726708093543.
DeHart-Davis, Leisha, and Sanjay K. Pandey. 2005. “Red Tape and Public Employees:: Does Perceived Rule Dysfunction Alienate Managers?” *Journal of Public Administration Research and Theory* 15(1):133–48. doi:10.1093/jopart/mui007.
DeHart-Davis, Leisha. 2009. “Green Tape and Public Employee Rule Abidance: Why Organizational Rule Attributes Matter.” *Public Administration Review* 69(5):901–10. doi:10.1111/j.1540-6210.2009.02039.x.
DeHart-Davis, Leisha. 2017. *Creating Effective Rules in Public Sector Organizations*. Washington: Georgetown University Press.
Dewe, Philip, Michael P. O'Driscoll, and Cary L. Cooper. 2010. *Coping with Work Stress: A Review and Critique*. Chichester, Malden: Wiley-Blackwell.
Erat, Sanjiv, and Uri Gneezy. 2012. “White Lies.” *Management Science* 58(4):723–33. doi:10.1287/mnsc.1110.1449.
Eterno, Eli, and John Silverman. 2012. *The Crime Numbers Game: Management by Manipulation*. Boca Raton: CRC Press.
Fausset, Richard, and Alan Blinder. 2015. “Atlanta School Workers Sentenced in Test Score Cheating Case.” Retrieved March 15, 2019 (https://www.nytimes.com/2015/04/15/us/atlanta-school-workers-sentenced-in-test-score-cheating-case.html).
Gerrish, Ed. 2017. “The Effect of the Child Support Performance and Incentive Act of 1998 on Rewarded and Unrewarded Performance Goals.” *Journal of Policy Analysis and Management* 36(1):65–96. doi:10.1002/pam.21957.

Gino, Francesca, Maurice E. Schweitzer, Nicole L. Mead, and Dan Ariely. 2011. "Unable to resist temptation: How self-control depletion promotes unethical behavior." *Organizational Behavior and Human Decision Processes* 115(2):191–203. doi:10.1016/j.obhdp.2011.03.001.
Glynn, Adam N. 2013. "What Can We Learn with Statistical Truth Serum?" *Public Opinion Quarterly* 77(S1):159–72. doi:10.1093/poq/nfs070.
Grant, Adam M., and Elizabeth M. Campbell. 2007. "Doing good, doing harm, being well and burning out: The interactions of perceived prosocial and antisocial impact in service work." *Journal of Occupational and Organizational Psychology* 80(4):665–91. doi:10.1348/096317906X169553.
Grant, Adam M., and David A. Hofmann. 2011. "Outsourcing inspiration: The performance effects of ideological messages from leaders and beneficiaries." *Organizational Behavior and Human Decision Processes* 116(2):173–87. doi:10.1016/j.obhdp.2011.06.005.
Grant, Adam M. 2012. "Leading with Meaning: Beneficiary Contact, Prosocial Impact, and the Performance Effects of Transformational Leadership." *Academy of Management Journal* 55(2):458–76. doi:10.5465/amj.2010.0588.
Heckman, James J., Carolyn J. Heinrich, Pascal Courty, Gerald Marschke, and Jeffrey Smith. 2011. *The Performance of Performance Standards:* W.E. Upjohn Institute.
Heinrich, Carolyn J., and Gerald Marschke. 2010. "Incentives and their dynamics in public sector performance management systems." *Journal of Policy Analysis and Management* 29(1):183–208. doi:10.1002/pam.20484.
Hill, Carolyn J., and Laurence E. Lynn. 2015. *Public management: Thinking and acting in three dimensions*. 2nd ed. Thousand Oaks, California: CQ Press.
Hood, Christopher. 2006. "Gaming in Targetworld: The Targets Approach to Managing British Public Services." *Public Administration Review* 66(4):515–21. doi:10.1111/j.1540-6210.2006.00612.x.
Hubbard, Michael L., Rachel A. Casper, and Judith T. Lessler. 1989. "Respondents' reactions to item count lists and randomized response." *Proceedings of the Survey Research Section of the American Statistical Association* 544-548.
Huff, C., and Dusting Tingley. 2015. "Who are these people?: Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2(3). doi:10.1177/2053168015604648.
Imai, Kosuke. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106(494):407–16. doi:10.1198/jasa.2011.ap10415.
Jacob, B. A., and S. D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118(3):843–77. doi:10.1162/00335530360698441.
Jakobsen, Morten, and Rasmus Jensen. 2015. "Common Method Bias in Public Management Studies." *International Public Management Journal* 18(1):3–30. doi:10.1080/10967494.2014.997906.
Kelman, S., and J. N. Friedman. 2009. "Performance Improvement and Performance Dysfunction: an Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service." *Journal of Public Administration Research and Theory* 19(4):917–46. doi:10.1093/jopart/mun028.
Kish-Gephart, Jennifer J., David A. Harrison, and Linda K. Treviño. 2010. "Bad Apples, Bad Cases, and Bad Barrels: Meta-Analytic Evidence About Sources of Unethical Decisions at Work." *Journal of Applied Psychology* 95(1):1–31. doi:10.1037/a0017103.

Koning, Pierre, and Carolyn J. Heinrich. 2013. "Cream-Skimming, Parking and Other Intended and Unintended Effects of High-Powered, Performance-Based Contracts." *Journal of Policy Analysis and Management* 32(3):461–83. doi:10.1002/pam.21695.

Kouchaki, Maryam, and Sreedhari D. Desai. 2015. "Anxious, Threatened, and Also Unethical: How Anxiety Makes Individuals Feel Threatened and Commit Unethical Acts." *Journal of Applied Psychology* 100(2):360–75. doi:10.1037/a0037796.

Kroll, Alexander, and Dominik Vogel. 2014. "The PSM-Leadership Fit: A Model of Performance Information Use." *Public Administration* 92(4):974–91. doi:10.1111/padm.12014.

Li, Jiayuan. 2015. "The paradox of performance regimes: Strategic responses to target regimes in Chinese local government." *Public Administration* 93(4):1152–67. doi:10.1111/padm.12197.

Moynihan, Donald P., and Sanjay K. Pandey. 2010. "The Big Question for Performance Management: Why Do Managers Use Performance Information?" *Journal of Public Administration Research and Theory* 20(4):849–66. doi:10.1093/jopart/muq004.

Moynihan, Donald P., Sanjay K. Pandey, and Bradley E. Wright. 2012. "Prosocial Values and Performance Management Theory: Linking Perceived Social Impact and Performance Information Use." *Governance* 25(3):463–83. doi:10.1111/j.1468-0491.2012.01583.x.

Nuno, Ana, and Freya A. St. John. 2015. "How to ask sensitive questions in conservation: A review of specialized questioning techniques." *Biological Conservation* 189:5–15. doi:10.1016/j.biocon.2014.09.047.

Olsen, Asmus L., Frederik Hjorth, Nikolaj Harmon, and Sebastian Barfort. 2019. "Behavioral Dishonesty in the Public Sector." *Journal of Public Administration Research and Theory* 29(4):572–90. doi:10.1093/jopart/muy058.

Pandey, Sanjay K., and Hal G. Rainey. 2007. "Public Managers' Perceptions of Organizational Goal Ambiguity: Analyzing Alternative Models." *International Public Management Journal* 9(2):85–112. doi:10.1080/10967490600766953.

Parker, Donald F., and Thomas A. DeCotiis. 1983. "Organizational determinants of job stress." *Organizational Behavior and Human Performance* 32(2):160–77. doi:10.1016/0030-5073(83)90145-9.

Portillo, Shannon, and Leisha DeHart-Davis. 2009. "Gender and Organizational Rule Abidance." *Public Administration Review* 69(2):339–47. doi:10.1111/j.1540-6210.2008.01978.x.

Radin, Beryl A. 2006. *Challenging the performance movement: Accountability, complexity, and democratic values*. Washington, D.C: Georgetown University Press.

Rainey, Hal G., Sanjay K. Pandey, and Barry Bozeman. 1995. "Research Note: Public and Private Managers' Perceptions of Red Tape." *Public Administration Review* 55(6):567–74. doi:10.2307/3110348.

Schott, Carina, and Adrian Ritz. 2018. "The Dark Sides of Public Service Motivation: A Multi-level Theoretical Framework." *Perspectives on Public Management and Governance* 1(1):29–42. doi:10.1093/ppmgov/gvx011.

Selart, Marcus, and Svein T. Johansen. 2011. "Ethical Decision Making in Organizations: The Role of Leadership Stress." *Journal of Business Ethics* 99(2):129–43. doi:10.1007/s10551-010-0649-0.

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal. 2015. "Self-Serving Justifications." *Current Directions in Psychological Science* 24(2):125–30. doi:10.1177/0963721414553264.

Siegel, Jason T., and Mario Navarro. 2019. “A conceptual replication examining the risk of overtly listing eligibility criteria on Amazon’s Mechanical Turk.” *Journal of Applied Social Psychology* 49(4):239–48. doi:10.1111/jasp.12580.
Stritch, Justin M., Mogens J. Pedersen, and Gabel Taggart. 2017. “The Opportunities and Limitations of Using Mechanical Turk (MTurk) in Public Administration and Management Scholarship.” *International Public Management Journal* 20(3):489–511. doi:10.1080/10967494.2016.1276493.
Tsai, Chi-lin. 2019. “Statistical analysis of the item-count technique using Stata.” *The Stata Journal: Promoting communications on statistics and Stata* 19(2):390–434. doi:10.1177/1536867X19854018.
Yam, Kai C., and Scott J. Reynolds. 2016. “The Effects of Victim Anonymity on Unethical Behavior.” *Journal of Business Ethics* 136(1):13–22. doi:10.1007/s10551-014-2367-5.
Yang, Liu-Qin, Hongsheng Che, and Paul E. Spector. 2008. “Job stress and well-being: An examination from the view of person-environment fit.” *Journal of Occupational and Organizational Psychology* 81(3):567–87. doi:10.1348/096317907X243324.
Zahra, Shaker A., Richard L. Priem, and Abdul A. Rasheed. 2007. “Understanding the Causes and Effects of Top Management Fraud.” *Organizational Dynamics* 36(2):122–39. doi:10.1016/j.orgdyn.2007.03.002.

Appendix A: Screening Survey

1. Which of the following labels characterizes the organization you work for best?
   - I don’t have a job
   - Private company or organization
   - Nonprofit organization
   - Public administration or government

2. At work, how much of your time do you spend on the following activities? Please allocate 100% of your work time.
   - Working with customers, clients, or citizens
   - Dealing with data and performance information
   - Supervising other people and making decisions
   - Planning and coordinating work with others
   - Working on cases or files by myself

Appendix B: Operationalizations

Variable	Operationalization
Performance Cheating	How many of the following statements do you agree with? We do not need to know which ones you would agree with, just how many. As a part of my job, I… • Make decisions and allocate funds based on performance data • Collect and consolidate performance data from colleagues or stakeholders • Explain performance data so my supervisor understands them • Correct errors and clean data but don't really make any decisions • [Alter data so that the results look better.]
Prosocial impact (Grant and Campbell 2007; Moynihan et al. 2012) (α = .93)	To what extent do you agree or disagree with the following statements? • I feel that my work makes a positive difference in other people's lives • I am very aware of the ways in which my work is benefiting others • I am very conscious of the positive impact my work has on others • I have a positive impact on others in my work on a regular basis (1 = not at all; 7 = very much)
Job-related anxiety (Parker and DeCotiis 1983) (α =.83)	To what extent do you agree or disagree with the following statements? • I have felt fidgety or nervous as a result of my job • My job gets to me more than it should • There are lots of times when my job drives me right up the wall • Sometimes when I think about my job I get a tight feeling in my chest • I feel guilty when I take time off from job (1 = not at all; 7 = very much)
Red tape (Rainey et al. 1995)	If red tape is defined as burdensome administrative rules and procedures that have negative effects on the organization's performance, please assess the level of red tape in your organization. (0 = lowest level; 10 = highest level)
Organizational performance focus (α = .77)	To what extent do you agree or disagree with the following statements? • Performance data are emphasized and discussed in documents and meetings • Good and bad performance yields different consequences for employees. • Appraisals are linked to goal achievement and performance data • My organization rewards employees for the achievement of performance targets (1 = not at all; 7 = very much)
Performance data exposure	At work, how much of your time do you spend on the following activities? • Dealing with data and performance information [*For other response categories, see the full wording of the screening survey in Appendix A*]
Age	How old are you? • 1 = 18 to 25 years • 2 = 26 to 35 years • 3 = 36 to 45 years

	• 4 = 46 to 55 years • 5 = Older than 55 years
Gender	What is your gender? • 0 = Man • 1 = Woman [Note: The survey contain a third category “Other” but none of the participants responded in this category]
Education	Which of the following options describes your level of educational attainment best? • 1 = High school degree; • 2 = Some college • 3 = 2-year associate degree • 4 = 4-year college degree • 5 = Graduate degree or higher

Appendix C: Full results table

Multivariate analysis of list experiment

Variable	Est. prop. of cheating		Multivariate			
	Est.	SE	Est.	SE	Est.	SE
Sensitive item						
Intercept	0.060	0.066	-14.964	3.350***		
Prosocial impact			1.070	0.569†		
Job-related anxiety			0.437	0.190*		
Red tape			0.313	0.136*		
Organizational performance focus			0.229	0.425		
Performance data exposure			0.003	0.023		
Age			0.089	0.037*		
Gender (1 = female)			1.246	0.704†		
Education			-0.740	0.247**		
Control items			$h_0(y; x, \psi_0)$		$h_1(y; x, \psi_1)$	
Intercept	1.946	0.044***	-0.340	0.313	7.199	3.514*
Prosocial impact			-0.062	0.033†	-0.374	0.520
Job-related anxiety			-0.032	0.026	0.249	0.138†
Red tape			0.015	0.015	-0.469	0.143**
Organizational performance focus			0.137	0.036***	0.014	0.266
Performance data exposure			0.006	0.002**	0.028	0.013*
Age			-0.010	0.004*	-0.084	0.025***
Gender (1 = female)			0.153	0.073*	-1.131	0.616†
Education			0.047	0.036	0.353	0.193†
n	964		964			

Note: Estimated coefficients from the item count technique regression models, where the sensitive item is whether or not participants 'alter data so that the results look better' in their job. Coefficients in the first column are based on linear least squares estimates. All other coefficients are based on logistic regression models with unconstrained maximum likelihood estimation. † p < .1, * p < .05, ** p < .01, p < .001

Appendix D: Check of assumptions

We assessed the applicability of the proposed statistical procedures by testing violations of the basic assumptions of list experiments (Blair and Imai 2012; Imai 2011). First, we did not find any indication that the sensitive item changed the response of the treatment group to the control items (i.e., design effects). The corresponding test (Blair and Imai 2012) of differences in response to the control items between treatment and control group failed to reject the null hypothesis of no design effect (Bonferroni corrected $p = 1.00$). Second, we tested if there are participants in the treatment group who lied about the sensitive item. Blair and Imai (2012) discuss two situations in which this could be the case: “ceiling effects” and “floor effects.” Ceiling effects occur if participants agree with all control items and the sensitive item but do not indicate that they agree with all items as this would reveal to the researchers that they agree with the sensitive item. Floor effects occur when participants only agree with the sensitive item but answer that they do not agree with any of the items. A histogram of the responses to the item list (see the figure below) shows that disagreement with all items is seldom with 29 cases (6.03 %) in the control group and 32 cases (6.63 %) in the treatment group. Agreement with all items in the treatment group is even less common. Only eight participants agree with all items (1.66 %). To rule out that ceiling or floor effects bias our estimations, we estimated the intercept-only model with the modeling adjustments suggested by Blair and Imai (2012). There are no differences between the unadjusted and the adjusted estimations. In all models, the intercept is 0.06. We, therefore, conclude that the assumptions for list experiments are met and our estimations are valid.

Fig. for Appendix D: Number of items participants agreed with per treatment group