Title: Negativity Bias, Gender, and Performance Information Use

Authors:
Sean Webeck
Jill Nicholson-Crotty
Sean Nicholson-Crotty

Abstract:

A key assumption of performance management is that individuals use performance information to make decisions. Very recent work in this area suggests that behavioral factors, such as cognitive biases, may help us understand this phenomenon. We leverage the concept of negativity bias in order to develop expectations about the relationship between gender and performance information use. Using an experimental design, we demonstrate that public sector workers exhibit a negativity bias when assessing performance information. We also find that females and males respond differently to negative information but that this also depends on what the bureaucrats were asked to assess.

Keywords: performance information, bureaucratic decision making, gender, framing, negativity bias

Governments throughout the world have invested heavily over the last 30 years developing systems to measure the performance of public enterprises. Nonetheless, one of the most persistent questions in the study of public programs and organizations is the degree to which managers utilize the information produced by these systems. Numerous studies have identified a long list of correlates, both at the individual and organizational level, that influence the use of performance information. Among these correlates, authors have investigated the impact of managerial gender, but this variable has often been added to analyses with little or no theoretical justification.

Building on recent research, which suggests that non-cognitive factors may be important predictors of performance information use, we argue in this paper that there are good reasons to expect gender to influence how public sector actors interpret performance information. This assertion is grounded primarily in a long line of psychological research which suggests differences in the decision-making processes of men and women, but also in findings from the public administration literature regarding differences between the genders in management style, representation behavior, and interaction with the organizational environment.

In this paper, we focus specifically on negativity bias in order to develop specific testable hypotheses regarding gender and performance information use. This well-established cognitive bias results in individuals responding more strongly to negative information than to comparable information that is framed in a positive way. While both men and women suffer from negativity bias, research suggests that women react more strongly to negatively framed information than do men. Additionally, research suggests the gender differences in negativity bias decrease or go away when the potential negative outcome directly affects the decision maker, rather that someone or something else. From these bodies of work, we draw three interrelated propositions considering the use of performance information. First, we expect that all bureaucrats will exhibit negativity bias when interpreting performance information. Second, we hypothesize that women respond more strongly to negatively valenced information than do men. And, finally, we offer the expectation that the gender differences in negativity bias change depending upon whether information being interpreted is about the individual or an organization with which they are affiliated.

To test for relationships between cognitive biases, gender, and the interpretation of performance information, we take an experimental approach. Specifically, we design a survey experiment in which we first randomly assign individuals to one of two performance vignettes: one in which we ask them to assess the performance of an unnamed city in which they are the city manager and one in which we ask them to assess their *personal* performance. Respondents in both groups are randomly shown negatively or positively valenced performance information about an equivalent level of citizen satisfaction with city services.

In line with our proposition, results suggest a clear negativity bias among individuals who work in or who have at least five years' experience working in the public sector. Also, consistent with our expectations, women responded more strongly to negative stimuli in the City vignette than men did. Finally, we find that the gender difference in negativity bias went away when decision makers were asked to rate their own performance, rather than that of their city. It is important to note that, while we use psychological research to suggest mechanisms which might cause these findings, we are *not* specifically testing (for) mechanisms in this piece.

In subsequent sections we provide a discussion of the use of performance information from public administration, including a recent movement to better incorporate a behavioral (i.e., psychological) perspective. We then turn to literature from psychology which allows us to build

our arguments about negativity bias, gender, and self-threat. We move on to a discussion of the experiment and the results before we conclude with a discussion of this research by casting it in the larger context for both behavioral public administration and the use of performance information by bureaucrats.

PERFORMANCE INFORMATION USE IN THE LITERATURE

An important question in the study of public administration pertains to why bureaucrats use the information created by nearly ubiquitous performance measurement and management systems (Moynihan and Pandey 2010). There has been a great deal of research on this topic. One outcome of this program has been that scholars now conceptualize three groups of 'users': bureaucrats, citizens, and political elites (Van Dooren and Van de Walle 2011). Because of the focus of the paper, this review will begin with a focus on the use of performance metrics by bureaucrats, but will return literature on other 'end users' when highlighting very recent work which takes a behavioral approach to information use. Behn (2003) argued that bureaucrats use performance metrics to evaluate, control, budget, motivate, promote, celebrate, learn, and improve. Yet, he emphasized that the last of these—improve—was the most important. In fact, he argued that all the rest 'are simply means for achieving this ultimate purpose', which pertains to improving performance. Taking a slightly different approach, rather than looking at specific actions, Moynihan (2010) suggests there are four strategies public managers can employ when using performance information: passive, political, perverse, and purposeful. Yet, for Moynihan, the existence of other interested actors (e.g., political principals and citizens) makes performance data more subjective than they are often argued to be.

Much of the empirical work on why and how bureaucrats choose to use that subjective information focuses on institutional factors. The findings in this research program point to a number of consistent predictors of information use, including: measurement system maturity (for examples see Berman and Wang 2000; Ho 2006; Taylor 2009), stakeholder involvement (for examples see Ho 2006; Bourdeaux and Chikoto 2008; Moynihan and Pandey 2010), leadership support (for examples see Moynihan and Ingraham 2004; Yang and Hsieh 2007; Moynihan and Lavertu 2012) (Moynihan and Ingraham 2004; Yang and Hsieh 2007; Moynihan and Lavertu 2012), support capacity (for examples see Berman and Wang 2000; Julnes and Holzer 2001; Moynihan and Hawes 2012), innovative culture (for examples see Moynihan 2005; Moynihan and Pandey 2010; Moynihan et al. 2012b), and goal clarity (for examples see Moynihan and Landuyt 2009; Moynihan et al. 2012b; Moynihan et al. 2012a).

This above classification scheme was developed in a recent and expansive review of the performance information use literature (see Kroll 2015). Interestingly, that review argues that despite all the work, the question of why bureaucrats use performance metrics remains 'under explored' (Kroll 2015). This may simply be an effect of previous research ignoring *individual-* level characteristics which might produce systematic variation in the use of performance information among bureaucrats (Kroll 2015; Moynihan et al. *forthcoming*). Those warnings not withstanding, a few *very* recent pieces have taken an individual-level behavioral approach to the use of performance information. A significant majority of these studies look at how citizens respond to performance information (Olsen 2013; Andersen and Hjortskov 2015; Olsen 2015a; Baekgaard and Serritzlew 2016; Barrows et al. 2016; Hvidman and Andersen 2016). But other work has looked at how politicians (Olsen 2014; Nielsen and Baekgaard 2015; George et al. 2016; Nielsen and Moynihan 2016) and bureaucrats (Moynihan 2008; Salge 2011; Nielsen 2013;

Moynihan 2015; Andersen and Moynihan 2016), use performance information. A key takeaway from these studies is that, depending on the circumstances, individuals exhibit a variety of cognitive biases and utilize a number of heuristics when responding to (ostensibly objective) performance information.

<div align="center">Gender and Performance Information Use: A Behavioral Approach</div>

Only a few studies tested for gender differences in the use of performance information (Kroll 2015). Unfortunately, the sex of the respondent is typically just included as a control variable in these studies, with little theoretical jsutification. In one of these, the variable had a positive and significant impact on information use (Taylor 2011). In another, the indicator of sex was positive but not significant in the authors' model of the use of externally generated information (Hammerschmid et al. 2013). In a final study, an indicator of gender was positively correlated with information use, but was not significant in either simple bivariate correlations or in the SEM estimated by the authors (Moynihan et al. 2012b). In our mind, these results suggest the potential for a gender difference in the use of performance information. These findings, and others, led Kroll (2015) to conclude that 'simple sociodemographic characteristics do not seem able to pick up on variations in data use' (476). Despite this conclusion, he suggested that differences among potential information users might still provide fruitful research, if researchers ask the right questions.

We argue in this paper that gender is an individual characteristic that may systematically influence the how and why bureaucrats choose to use performance information. In part, we suggest this because gender is an important factor in several research programs within public administration and management. More importantly, those bodies of work take a socialization perspective to suggest that women may make different decisions than their male counterparts when managing or working in public programs and organizations. For example, research suggests that female managers are more likely to solicit a variety of perspectives from subordinates and stakeholders before making a decision (Eagly et al. 1992). Additionally, research suggests that men and women use different criteria and assess different evidence when evaluating the performance of others in an organization (Sinclair and Kunda 2000). Similarly, studies suggest the impact of gender congruence between leader and subordinates on decisions is different for men and women (Grissom et al. 2012). Finally, research in representative bureaucracy suggests that female bureaucrats may make decisions based on shared characteristics or experiences with clients in a way that male administrators do not (Keiser et al. 2002; Meier and Nicholson-Crotty 2006).

Despite work in public administration and management, our expectations about gender differences in performance information use are primarily grounded in the field of psychology. This research will be reviewed in more detail later in this section, but for now it is sufficient to note that scholars have consistently found differences in risk tolerance between the sexes (for a review see Neelakantan 2010). Similarly, research suggests that men and women approach information with different cognitive biases (for examples see Dahlbom et al. 2011; Stevens and Hamann 2012). In short, there may exist explanations beyond socialization for us to observe gender differences in various decision-making processes.

Decision Making and Public Administration

Herbert Simon (1947) argued that decision making was essential to the practice and study of administrative organizations and that decisions should be the primary unit of analysis in decision-making research. Recent scholarship suggests the field could do more to embrace Simon's perspective (Meier 2015; Olsen 2015c). Other scholars argue the way forward on some of the field's long-standing problems is to adopt a behavioral perspective (Grimmelikhuijsen et al. 2016). Thus, by adopting some of the insights from psychology it might be possible to better understand how bureaucrats make decisions.

Building on this movement, this study takes an experimental approach to understanding the interplay between gender and cognitive biases in the use of performance information. Building off of recent work which suggests that cognitive bias influences the ways in which decision makers interpret such information (Nielsen 2013; Olsen 2013; Olsen 2015a; Olsen 2015b) , we turn to research from the field of psychology to provide insight into specific ways in which biases might influence the use of performance information differently between men and women.

Negativity bias, gender, and the use of performance information.

Scholars have long recognized that human beings have a 'negativity bias,' which simply means that we tend to respond more strongly to negative information than to comparable information that is framed in a positive way (Ito et al. 1998). A recent study from Denmark provides evidence for a negativity bias in the use of performance information. Olsen (2015a) shows that citizens evaluate public organizations differently if they believe the organization has a 90% satisfaction rate or a 10% dissatisfaction rate, clear evidence of a negativity bias. Olsen's work looks at citizen evaluations, but it might be that bureaucrats, individuals more accustomed to working with and thinking about performance information, will react more rationally to the information created by performance management systems. That is, we might expect bureaucrats, as experts, to *not* suffer from this same bias. At the same time, the 'universality' of the negativity bias suggests we should see this bias in bureaucrats (Ito et al. 1998; Ito and Cacioppo 2005). A lack of research on this question provides one motivation for this study. In line with the universality argument and Olsen's work, we expect bureaucrats to suffer from a negativity bias as well. More specifically, we expect bureaucrats will provide lower interpretation scores for performance information that is framed negatively compared to when it is framed positively, even when the metrics being compared are qualitatively similar. Stated formally,

H1: Public sector employees will indicate lower levels of performance when presented with negatively (rather than positively) framed performance metrics.

But, while the direction of this effect is largely consistent across people, we see individual differences in the magnitude of the negativity bias (Ito and Cacioppo 2005; Norris et al. 2011). Additionally, evidence suggests it may be fruitful for researchers to explore the role that gender may play in affective bureaucratic decision making.

According to the Evaluative Space Model (ESM), feeling 'good' and feeling 'bad' do not exist on a single continuum (Norris et al. 2011). Thus, while these are both emotional outcomes, the brain undergoes different processes of evaluation to arrive at these feelings. The ESM framework suggests humans process external stimuli according to the appetitive/approach system or the aversive/avoidance system (Bradley et al. 2001a; Ashare et al. 2013). The ESM and an understanding of the underlying structure of the brain helps psychologists to study potential

4

causes and consequences of these emotional systems. One line of inquiry growing out of this work suggests gender as a potential cause of variation in our emotional responses (Kret and De Gelder 2012). For example, numerous studies demonstrate women respond more strongly than men to negative stimuli (Bradley et al. 2001b; Whittle et al. 2011; Groen et al. 2013; Syrjänen and Wiens 2013).

What explains these differences? Some evidence suggests different brain functions might influence gender differences in negativity bias (Stevens and Hamann 2012). Of these, two stand out. One example is serotonin, which leads to differences in affective evaluation between males and females (Ashare et al. 2013). Another is regional brain activation. Specifically, amygdala activation plays a major role in gender differences regarding negativity bias (Stevens and Hamann 2012). At this point we want to reinforce to the reader that although we lay out a story in which physiological differences potentially influence the use of performance information differently between the two genders but that *we do not test for any mechanisms in this study* (e.g., negativity bias, self-threat bias, of gender differences). Despite this caveat, we believe that any evidence of a negativity bias will be stronger in females than males. Stated formally,

H2: The negativity bias will be more pronounced for females than for males.

Self-threat bias in the use of performance information.
Research suggests we also need to consider the self-serving bias when speculating about the relationship between gender, negativity bias, and the interpretation of performance information. Self-serving bias occurs when we view ourselves more favorably than other people (Roese and Olson 2007). Additionally, this bias allows us to take responsibility for positive behavioral outcomes but to deny that same responsibility when those behavioral outcomes are negative (Bradley 1978). This bias takes many forms, one of which is the self-serving attributional bias which suggests people attribute blame and praise to themselves differently than they do to others (Roese and Olson 2007). That is, it's not just that we attribute responsibility differently for ourselves depending upon outcomes but the way we attribute responsibility for ourselves is different than the way we attribute responsibility for other people.

The specific form of the self-serving bias that is more relevant here, however, is known as 'self-threat.' For this condition to occur, an individual must perceive the presence of a condition unfavorable to the self (Campbell and Sedikides 1999). This activates the self-serving judgment, which has the potential to influence affective regulation (Roese and Olson 2007). Because seeing yourself positively feels good (and feels better than seeing yourself negatively) this process may 'be used strategically to regulate affect (i.e., to restore the individual to an affective set-point following unpleasant experiences)' (Rose and Olson 2007,127). They argue that self-threat magnifies self-serving biases. This is because negative stimuli (i.e., threat) activate self-serving judgments more strongly than do positive stimuli. Additionally, self-threat bias leads people to process negative stimuli differently if the stimuli in question is directed at them than if it is directed at someone, or something, else. And, because men and women process negative stimuli differently, we expect the self-threat bias to influence the *negativity bias* differently between men and women. Thus, we expect a smaller gender difference in the negativity bias when respondents themselves are under threat than if some other person or entity is under threat. Stated formally,

H3: Gender differences in the negativity bias of performance assessments should be less pronounced when subjects are asked to assess their own performance compared to when they are asked to assess the performance of an unnamed city

It is worth noting that individuals in our experiments were not even asked to rate their own performance. Rather, we asked them to rate their performance if they had been working as a city manager of the unnamed city. We include the following excerpt from the survey to represent what respondents in the Self category saw: '*Assuming this is the only piece of performance data available to you, use the sliding scale (0-100) to assess YOUR overall performance over the last year:*'. Thus, due to the nature of the way we designed our survey, any degree of self-threat would be perceived rather than actual self-threat—respondents aren't even evaluating their own performance. For this reason, we believe that any evidence in support of *H3* would be meaningful and potentially larger, substantively, if respondents were to assess their own performance (e.g., in the field) after being given negatively valenced performance information.

DATA AND METHODS

Data Collection and Survey Instruments

We used surveys to collect data in two phases. Individuals were paid for their participation in both phases. We designed data collection instruments for both phases using Qualtrics. We utilized TurkPrime (www.turkprime.com) as a third-party platform to collect data from Workers on Amazon's Mechanical Turk (MTurk). TurkPrime offers researchers both greater flexibility and control over the design and implementation of online, crowdsourced research. Specifically, it was designed 'to optimize MTurk functionality for the needs of researchers' (Litman et al. 2016).

In the first phase, we ran a short survey that allowed us to screen respondents in two ways. First, respondents were asked to select the sector that best described their primary employment. Possible responses included: private for-profit, private not-for-profit, public, and N/A (e.g., unemployed, out of the workforce, etc.). We provided representative examples in case the sector type would confuse anyone. In addition to this question we then asked individuals if they had ever worked in each of the three sectors. Respondents could select 'yes' or 'no'. If they selected 'yes', respondents then saw an additional question in which they provided a numerical response for the number of years they worked in the respective sector. 1476 unique individuals completed the screening phase. Individuals passed as preliminarily qualified if they indicated they currently worked in the public sector or that they had at least five years of work experience in the public sector. Of these, individuals were disqualified for the following reasons: beeline responses (e.g., people indicated they had worked five years in each sector), 50 or more years of experience in any sector, 60 or more years of combined experience, and anyone who first indicated they worked in the public sector but then later indicated they had never worked in the public sector. This gave us 281 individuals who met our qualifications.

We used TurkPrime to send a message to the 281 qualified respondents about a special survey, inviting those individuals to take the second phase of the project. A series of experiments were embedded in that survey, though we only describe and report results from negativity bias experiments here. Individuals earned $1.85 for completing the second phase. Out of these

invitations for the second phase of the project 242 individuals began the survey. We received 222 responses and 206 could be considered complete.

The data from this experiment comes from a pilot study which is registered with Evidence in Governance and Politics (EGAP) under the following ID: 20160730AA. That instrument includes the test for the negativity bias.

Experiment Description

To test for negativity bias we ask respondents to respond to randomly generated performance information about a city in which they were the city manager.[1] Specifically, respondents were given a prompt about citizen satisfaction with strategic asset investment in an unnamed city. The wording (see below) of the performance information metric comes from KCStat, the performance management system dashboard for the city of Kansas City, Missouri. The performance levels are equivalent, but in the negative frame, respondents are told about the percent of citizens that are dissatisfied, while in the positive frame their attention is drawn to the number of citizens that are satisfied with the city's performance.   Respondents were also randomly assigned to one of two vignettes (City or Self), where one group rated the performance of the *city* if they were the city manager, while the other was asked to rate their *own* performance had they been the city manager. The only differences which existed in the wording respondents saw reflected whether they should rate the city or themselves.

To provide a representative example of what respondents saw, we include two examples from the City vignette. The first example shows the prompt that individuals saw. Notice 'THE CITY' is in all caps. Individuals in the Self vignette saw 'YOU' and 'YOUR' in capital letters. We did this to increase the chances that individuals were responding to the prompt in the way we wanted them to.

> *For the next few questions, imagine that you are a city manager of a city in the United States of America. Your city just released its yearly performance metrics and the mayor wants to know how you think THE CITY performed over the course of the last year. In the following question you will be given performance information. You will then be asked to provide an assessment of THE CITY's performance over the last year given this performance information.*

An example of the negative frame follows:

> *Strategic Asset Investment: ##% of citizens were dissatisfied with the way the city maintained city streets, sidewalks, and other infrastructure.   Assuming this is the only piece of performance*

---

[1] We based the design of our experiment on previous work in this area Olsen, A. L. 2013. 'Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information.' *Judgment and Decision Making* 8(3): 365.
, Olsen, A. L. 2015a. 'Citizen (dis) satisfaction: An experimental equivalence framing study.' *Public Administration Review* 75(3): 469-478..

data available to you, use the sliding scale (0-100) to assess the
overall performance of THE CITY over the last year:

We additionally randomized positively and negatively valenced performance information. The range of positive metrics was 75-95% satisfied. The range of negative metrics was 5-25% dissatisfied. After seeing their prompt and performance metric, respondents then provided their performance assessment using a sliding scale. See Olsen (2015a) for a discussion of the reasoning behind setting up the performance metrics in this way. Tables 1 and 2 show the descriptive statistics for the actual raw performance metrics individuals saw based upon vignette and gender.

Before continuing we will use Table 1 to clarify how to read these Tables. Let's look at two of the four numbers from Table 1. Specifically, we'll look at 15.02 and 85.45. 15.02 represents the mean of the raw number seen by respondents in the city vignette who saw the negative frame. Remember, these respondents saw this raw number in the context of citizen dissatisfaction. The equivalent value would be 84.98 percent of citizens were satisfied with the service in question. When we move to the positive frame, the mean value seen by respondents was 85.45 (percent satisfied). From these values, we know that randomization was successful in creating statistically equivalent positive and negative information presentations across our key indicators of interest, including respondent gender and city or self vignette assignment.

<<<Insert Table 1 about here>>>

<<<Insert Table 2 about here>>>

RESULTS

This section presents the findings from the experiments described above. As a reminder, our expectations were that bureaucrats would manifest the negativity bias, responding to negatively valenced performance information more strongly than positively valenced information. (H1). We also expect that the negativity bias will appear more strongly in women than in men (H2), and that the effects of the negativity bias would diminish when respondents were asked to rate their own performance (self vignette) versus when they rated the city's performance (city vignette) (H3).

Negativity Bias

We begin with Table 3, which provides the mean performance rating responses for the four vignette-frame groups (City, negative [51 respondents]; City, positive [49 respondents]; Self, negative [54 respondents]; and Self, positive [52 respondents]). As discussed above, the values respondents saw represented statistically *equivalent* performance information. Thus, in the absence of negativity bias, the performance assessments across the two vignettes should be statistically equivalent. As Table 3 clearly shows, they are not. But, in order to say conclusively whether or not the data support our first hypothesis, we need to do some further analysis.

<<<Insert Table 3 about here>>>

To examine, between group differences in the responses, we provide a kernel density plot as well as an analysis of variance test (ANOVA). Figure 1 (the kernel density plot) shows two distributions, with the positive frame represented by the thin black line and the negative frame represented by the thick gray line. These are distributions across the entire sample (i.e., in this plot we are not discriminating between City and Self or male and female). The negative frame (thick line) has a clear left skew with a shorter peak while the positive frame has a relatively normal distribution with a peak around the mean, which provides visual evidence for the negativity bias.

<<<Insert 'Figure 1 – Overall Negativity Bias Difference chart about here>>>

Further evidence of the negativity bias comes from our ANOVA test, where we included a Scheffé multiple-comparison test option in our analysis to highlight the between-group differences (e.g., City, negative vs. positive frames; Self, negative vs. positive frames). Table 4 shows these between-group comparisons. The two highlighted cells represent our relationships of interest. Both suggest assessments are meaningfully lower when information is negatively rather than positively framed and both are significant at the $p=.05$ level. Taken together, the evidence supports our first hypothesis. Bureaucrats appear to manifest a negativity bias when using performance information.

<<<Insert Table 4 about here>>>

Gender

We now turn to our attention to the tests of our two remaining hypotheses. Table 5, breaks mean performance ratings for positive and negative information groups down by gender, as well as by self vs. city treatment. As we showed in Table 2 there were no significant differences in terms of the raw performance metrics seen by individuals in these various groups.

We can begin by examining the first two rows of Table 5, which present male and female responses to negatively and positively valenced performance information in the evaluation of the city. If the negativity bias were the same for men and women, we would expect the same within group impact size. What we see is that male respondents which received negative information gave a subjective performance assessment 10.85 points lower (85.85-75) than the *equivalent* raw performance score which they were given. Women, on the other hand, rated their city 17.76 points lower (84.08-66.32) than the *equivalent* raw measure of citizen satisfaction they saw . The positive information groups for both men and women also gave subjective assessments that were lower than the presented metric, though again the differences were much smaller (5.05 and 1.75). The drop in subjective assessment for women presented with negative versus positive

information appears to be substantively larger than the one for men, which would suggest that the negativity bias is stronger among women, but we again need the Scheffé test to accurately gauge these between group differences.

<<<Insert Table 5 about here>>>

Staying focused on hypothesis 2 and gender differences in the *city* evaluations, that test is presented in Table 6 and confirms the results in Table 5. For males, in the difference in response ratings between those in the negative and positive frames falls below the threshold for statistical significance in this more rigorous test. Comparatively, the large difference in the negative versus positive group responses for females assessing city performance remains significant at the $p=.01$ level. Taken together, these results support our expectation that women will respond more strongly to negative performance information than will men.

<<<Insert Table 6 about here>>>

The self-assessments by men and women, presented in the last 2 rows of Table 5, allows us to examine our third hypothesis, which suggests that self-threat should diminish differences in the negativity bias between genders. The findings suggest that men and women responded equivalently to negative information when asked to assess their personal performance, giving subjective assessments that were 14.36 and 14.31 lower, respectively, than the raw performance information they were provided. The men and women given positively valenced information both provided assessments that were very similar to the raw performance metric. The drops in performance assessments between men and women do not appear to be distinct from one another and darkened cells in the bottom two rows of the AVOVA in Table 6 confirm that the difference between these values is not statistically significant.

As a final assessment of these results, Figure 2 shows a kernel density plot showing the four groups which saw the negative frame across the two vignettes. In this Figure, we represent the respondents from the City vignette with a thin line and respondents from the Self vignette with a thick line. Visually, a couple of things stand out. First, the difference in the City vignette seems quite large around the peak(s) of both distributions. Second, female responses in the City vignette have a strong left skew. Finally, these differences between the two groups (largely) go away when we look at responses from the Self vignette. Figure 2 provides additional evidence for both *H2* and *H3*. It seems that females respond more strongly to negative performance information than men when asked to assess an organization (for which they might work) but that this difference goes away when females and males are asked to rate their own performance.

DISCUSSION AND CONCLUSIONS

10

Despite all the work on performance measurement and management, a prominent scholar recently commented that 'we know little about the basic tendency of individuals to incorporate and use performance information' (Moynihan 2015, 33). In part this lack of understanding may be due to a long-standing assumption in the literature that decisions about the use performance information are driven primarily by cognitive factors. Recent work has challenged this assumption and demonstrated that noncognitive processes, such as bias, have a profound influence on the ways in which human beings utilize information about the outputs and outcomes produced by public organizations. This study makes a significant contribution to this growing body of literature.

Previous research suggests citizens are prone to exhibit a negativity bias when confronted with performance information (Olsen 2015a). In this study we provided evidence that people with public sector work experience (i.e., bureaucrats) exhibit this bias as well. We also demonstrate that the negativity bias is stronger in women than in men, when assessing performance of their organization, but that the gender differences disappear when subjects are evaluating their *personal* performance. In the remainder of this section we discuss some of the possible implications for these findings. We also discuss the differing connotations for the term 'bias' across fields and the ways in which these have hindered research in this area. Finally, we note some directions for future research.

The finding of negativity bias in the interpretation of performance among bureaucratic actors is significant for several reasons. These findings suggest that the valence (positive/negative) of performance information has a substantial influence on the ways in which bureaucrats respond to that information, even when the underlying level of performance is exactly the same. This implies that performance measurement systems which emphasize positive factors such as 'satisfaction' or 'percent goals met' may induce a smaller response from bureaucrats tasked with utilizing information than will systems that emphasize negative outcomes such as 'dissatisfaction.' This is significant because performance management philosophies based on expectancy or motivation theory, along with those coupled with transformational leadership styles, are more likely to highlight positive rather than negative valence. This is not imply that the negativity bias may not induce an *overreaction* to negatively valenced information, causing managers and bureaucrats to over correct, but an awareness of the different reactions that differently framed information can have is important to understanding the use of that information.

The finding that this negativity bias manifests differently in men and women also has potentially significant implications for the efficacy of public performance management systems. Our results imply that the same level of organizational performance may induce a different reaction in male and female bureaucrats and public managers depending on how it is framed. Specifically, the implication is that negatively valenced information will produce a stronger reaction in women than in men when they are assessing the organization, while positively framed information will have a roughly equivalent impact on perceptions across the genders. This means that information from the same performance management system could be interpreted and used differently by male and female employees and managers. Again, this recognition does not lead to an obvious conclusion about the ideal design of these systems, but it does suggest that gender difference in the use of performance information is a factor that should be considered.

One potentially confounding aspect of this study which we did not explore is stereotype threat. Stereotype threat is considered a self-evaluative threat in which a member of a group about which a negative stereotype exists begins to fulfill the negative aspects of the stereotype,

even if not explicitly prompted (Steele and Aronson 1995). Previous research demonstrates this applies to females when asked to perform mathematical problems (Spencer et al. 1999). It is possible that even though unprompted about any gender differences that females in this study experienced the stereotype bias. Within the context of the current study it is impossible to draw a conclusive response on this question because of the study's design. That being said, females had higher average responses in three of the four vignette-frame groups. The only one in which males had a higher average score was the City-negative group and we argue that Hypothesis 2 explains this finding. In short, further research could explore stereotype threat and the use of performance information but we are not overly concerned with this bias in the present study.

We believe the findings from this study reinforce the value of research on non-cognitive processes in public sector decision making. To fully realize that potential, however, it is important to distinguish the different conceptions of the term 'bias.' When psychologists and other behavioral researchers discuss 'bias' they mean to describe the functions in the brain and processing of information that leads individuals to make sub-optimal decisions; sub-optimal in this sense is in line with the idea of maximization in expected utility theory (Von Neumann and Morgenstern 1953). In this view, bias is simply a way to describe the way that humans process information. In public administration, on the other hand, the idea of bias tends to be highly normative and often something to be avoided. We can see this definition at work in many of the foundational concepts of the field including neutral competence (Wilson 1887), bounded rationality (Simon 1946), and representative bureaucracy (Kingsley 1944).

Recent research suggests that significant gains in our understanding of public decision making can be made if scholars in the field of public administration will embrace the neutral definition of 'bias' utilized in other disciplines. More than 60 years ago, Simon (1946) proposed that students of public administration should do a better job of thinking about and incorporating decision making into the study of the field. However, a recent article suggested this had 'not become the predominant approach in' the field (Meier 2015, 18). One reason scholars of public administration and public management have not made decision making the fundamental approach to our field may be due to the challenges which arise from thinking about and confronting the idea of cognitive bias. We can conceptualize bureaucrats exclusively as agents who respond in rational ways within institutional environments, but this approach might well miss key behavioral aspects that are essential to understanding public decision making.

REFERENCES

Andersen, S. C. and M. Hjortskov 2015. 'Cognitive Biases in Performance Evaluations.' *Journal of Public Administration Research and Theory* 26(4): 647-662.

Andersen, S. C. and D. P. Moynihan 2016. 'How Leaders Respond to Diversity: The Moderating Role of Organizational Culture on Performance Information Use.' *Journal of Public Administration Research and Theory*: 1-13.

Ashare, R. L., C. J. Norris, E. P. Wileyto, J. T. Cacioppo and A. A. Strasser 2013. 'Individual differences in positivity offset and negativity bias: Gender-specific associations with two serotonin receptor genes.' *Personality and Individual Differences* 55(5): 469-473.

Baekgaard, M. and S. Serritzlew 2016. 'Interpreting performance information: Motivated reasoning or unbiased comprehension.' *Public Administration Review* 76(1): 73-82.

Barrows, S., M. Henderson, P. E. Peterson and M. R. West 2016. 'Relative Performance Information and Perceptions of Public Service Quality: Evidence From American School Districts.' *Journal of Public Administration Research and Theory*: muw028.

Behn, R. D. 2003. 'Why measure performance? Different purposes require different measures.' *Public Administration Review* 63(5): 586-606.

Berman, E. and X. Wang 2000. 'Performance measurement in US counties: Capacity for reform.' *Public Administration Review* 60(5): 409-420.

Bourdeaux, C. and G. Chikoto 2008. 'Legislative influences on performance management reform.' *Public Administration Review* 68(2): 253-265.

Bradley, G. W. 1978. 'Self-serving biases in the attribution process: A reexamination of the fact or fiction question.' *Journal of personality and social psychology* 36(1): 56.

Bradley, M. M., M. Codispoti, B. N. Cuthbert and P. J. Lang 2001a. 'Emotion and Motivation I: Defensive and appetitive reactions in picture processing.' *Emotion* 1(3): 276-298.

Bradley, M. M., M. Codispoti, D. Sabatinelli and P. J. Lang 2001b. 'Emotion and Motivation II: Sex differences in picture processing.' *Emotion* 1(3): 300-319.

Campbell, W. K. and C. Sedikides 1999. 'Self-threat magnifies the self-serving bias: A meta-analytic integration.' *Review of general Psychology* 3(1): 23.

Dahlbom, L., A. Jakobsson, N. Jakobsson and A. Kotsadam 2011. 'Gender and overconfidence: are girls really overconfident?' *Applied Economics Letters* 18(4): 325-327.

Eagly, A. H., S. J. Karau and B. T. Johnson 1992. 'Gender and leadership style among school principals: A meta-analysis.' *Educational Administration Quarterly* 28(1): 76-102.

George, B., S. Desmidt, P. Nielsen and M. Baekgaard 2016. Rational planning and politicians' attitudes to spending and reform: replication and extension of a survey experiment. 2016 IRSPM Annual Conference.

Grimmelikhuijsen, S., S. Jilke, A. Leth Olsen and L. Tummers 2016. 'Behavioral Public Administration: Combining Insights from Public Administration and Psychology.' *Public Administration Review*.

Grissom, J. A., J. Nicholson-Crotty and L. Keiser 2012. 'Does my boss's gender matter? Explaining job satisfaction and employee turnover in the public sector.' *Journal of Public Administration Research and Theory*.

Groen, Y., A. Wijers, O. Tucha and M. Althaus 2013. 'Are there sex differences in ERPs related to processing empathy-evoking pictures?' *Neuropsychologia* 51(1): 142-155.

Hammerschmid, G., S. Van de Walle and V. Stimac 2013. 'Internal and external use of performance information in public organizations: results from an international survey.' *Public Money & Management* 33(4): 261-268.

Ho, A. T.-K. 2006. 'Accounting for the value of performance measurement from the perspective of Midwestern mayors.' *Journal of Public Administration Research and Theory* 16(2): 217-237.

Hvidman, U. and S. C. Andersen 2016. 'Perceptions of Public and Private Performance: Evidence from a Survey Experiment.' *Public Administration Review* 76(1): 111-120.

Ito, T. and J. Cacioppo 2005. 'Variations on a human universal: Individual differences in positivity offset and negativity bias.' *Cognition & Emotion* 19(1): 1-26.

Ito, T. A., J. T. Larsen, N. K. Smith and J. T. Cacioppo 1998. 'Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations.' *Journal of personality and social psychology* 75(4): 887.

Julnes, P. d. L. and M. Holzer 2001. 'Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation.' *Public Administration Review* 61(6): 693-708.

Keiser, L. R., V. M. Wilkins, K. J. Meier and C. A. Holland 2002. 'Lipstick and Logarithms: Gender, institutional context, and representative bureaucracy.' *American Political Science Review* 96(3): 553-564.

Kingsley, J. D. 1944. *Representative Bureaucracy*. Yellow Springs, OH, Antioch Press.

Kret, M. E. and B. De Gelder 2012. 'A review on sex differences in processing emotional signals.' *Neuropsychologia* 50(7): 1211-1221.

Kroll, A. 2015. 'Drivers of performance information use: systematic literature review and directions for future research.' *Public Performance & Management Review* 38(3): 459-486.

Litman, L., J. Robinson and T. Abberbock 2016. 'TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences.' *Behavior research methods*: 1-10.

Meier, K. J. 2015. 'Proverbs and the Evolution of Public Administration.' *Public Administration Review* 75(1): 15-24.

Meier, K. J. and J. Nicholson-Crotty 2006. 'Gender, representative bureaucracy, and law enforcement: The case of sexual assault.' *Public Administration Review* 66(6): 850-860.

Moynihan, D. P. 2005. 'Goal-based learning and the future of performance management.' *Public Administration Review* 65(2): 203-216.

Moynihan, D. P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC, Georgetown University Press.

Moynihan, D. P. 2010. The Promises and Paradoxes of Performance-Based Bureaucracy. *The Oxford Handbook of American Bureaucracy*. R. F. Durant. New York, Oxford University Press.

Moynihan, D. P. 2015. 'Uncovering the Circumstances of Performance Information Use Findings from an Experiment.' *Public Performance & Management Review* 39(1): 33-57.

Moynihan, D. P. and D. P. Hawes 2012. 'Responsiveness to reform values: The influence of the environment on performance information use.' *Public Administration Review* 72(s1): S95-S105.

Moynihan, D. P. and P. W. Ingraham 2004. 'Integrative Leadership in the Public Sector A Model of Performance-Information Use.' *Administration & society* 36(4): 427-453.

Moynihan, D. P., A. Kroll and P. A. Nielsen *forthcoming*. Managerial Use of Performance Data by Bureaucrats and Politicians. *Experiments in Public Management Research: Challenges and Contributions*. O. James, S. Jilke and G. V. Ryzin, Cambridge University Press.

Moynihan, D. P. and N. Landuyt 2009. 'How do public organizations learn? Bridging cultural and structural perspectives.' *Public Administration Review* 69(6): 1097-1105.

Moynihan, D. P. and S. Lavertu 2012. 'Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART.' *Public Administration Review* 72(4): 592-602.

Moynihan, D. P. and S. K. Pandey 2010. 'The big question for performance management: Why do managers use performance information?' *Journal of Public Administration Research and Theory*: muq004.

Moynihan, D. P., S. K. Pandey and B. E. Wright 2012a. 'Prosocial values and performance management theory: Linking perceived social impact and performance information use.' *Governance* 25(3): 463-483.

Moynihan, D. P., S. K. Pandey and B. E. Wright 2012b. 'Setting the table: How transformational leadership fosters performance information use.' *Journal of Public Administration Research and Theory* 22(1): 143-164.

Neelakantan, U. 2010. 'Estimation and impact of gender differences in risk tolerance.' *Economic inquiry* 48(1): 228-233.

Nielsen, P. A. 2013. 'Performance Management, Managerial Authority, and Public Service Performance.' *Journal of Public Administration Research and Theory* 24(2): 431-458.

Nielsen, P. A. and M. Baekgaard 2015. 'Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment.' *Journal of Public Administration Research and Theory* 25(2): 545-569.

Nielsen, P. A. and D. P. Moynihan 2016. 'How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy.' *Journal of Public Administration Research and Theory*.

Norris, C. J., J. T. Larsen, L. E. Crawford and J. T. Cacioppo 2011. 'Better (or worse) for some than others: Individual differences in the positivity offset and negativity bias.' *Journal of Research in Personality* 45(1): 100-111.

Olsen, A. L. 2013. 'Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information.' *Judgment and Decision Making* 8(3): 365.

Olsen, A. L. 2014. The Appropriate Response to Performance Information by Citizens: A Large-scale Experiment with Local Politicians.

Olsen, A. L. 2015a. 'Citizen (dis) satisfaction: An experimental equivalence framing study.' *Public Administration Review* 75(3): 469-478.

Olsen, A. L. 2015b. 'The Numerical Psychology of Performance Information: Implications for Citizens, Managers, and Policymakers.' *Public Performance & Management Review* 39(1): 100-115.

Olsen, A. L. 2015c. ''Simon Said,' We Didn't Jump.' *Public Administration Review* 75(2): 325-326.

Roese, N. J. and J. M. Olson 2007. 'Better, stronger, faster: Self-serving judgment, affect regulation, and the optimal vigilance hypothesis.' *Perspectives on Psychological Science* 2(2): 124-141.

Salge, T. O. 2011. 'A behavioral model of innovative search: Evidence from public hospital services.' *Journal of Public Administration Research and Theory* 21(1): 181-210.

Simon, H. 1947. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York, Macmillan.

Simon, H. A. 1946. 'The proverbs of administration.' *Public Administration Review* 6(1): 53-67.

Sinclair, L. and Z. Kunda 2000. 'Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me.' *Personality and social psychology bulletin* 26(11): 1329-1342.

Spencer, S. J., C. M. Steele and D. M. Quinn 1999. 'Stereotype threat and women's math performance.' *Journal of experimental social psychology* 35(1): 4-28.

Steele, C. M. and J. Aronson 1995. 'Stereotype threat and the intellectual test performance of African Americans.' *Journal of personality and social psychology* 69(5): 797.

Stevens, J. S. and S. Hamann 2012. 'Sex differences in brain activation to emotional stimuli: a meta-analysis of neuroimaging studies.' *Neuropsychologia* 50(7): 1578-1593.

Syrjänen, E. and S. Wiens 2013. 'Gender moderates valence effects on the late positive potential to emotional distracters.' *Neuroscience letters* 551: 89-93.

Taylor, J. 2009. 'Strengthening the link between performance measurement and decision making.' *Public Administration* 87(4): 853-871.

Taylor, J. 2011. 'Factors influencing the use of performance information for decision making in Australian state agencies.' *Public Administration* 89(4): 1316-1334.

Van Dooren, W. and S. Van de Walle, Eds. 2011. *Performance Information in the Public Sector: How it is Used*. Governance and Management Series. New York, Palgrave Macmillan.

Von Neumann, J. and O. Morgenstern 1953. *Theory of Games and Economic Behavior*. Princeton, NJ, Princeton University Press.

Whittle, S., M. Yücel, M. B. Yap and N. B. Allen 2011. 'Sex differences in the neural correlates of emotion: evidence from neuroimaging.' *Biological psychology* 87(3): 319-333.

Wilson, W. 1887. 'The study of administration.' *Political science quarterly* 2(2): 197-222.

Yang, K. and J. Y. Hsieh 2007. 'Managerial effectiveness of government performance measurement: testing a middle-range model.' *Public Administration Review* 67(5): 861-879.

Table 1 – Negativity Bias, Gender, and Performance Information Use

Table 1 - Avg Raw Numbers Seen		
	Frame	
Evaluating	Negative	Positive
City	15.02	85.45
Self	14.04	84.88

Table 2 – Negativity Bias, Gender, and Performance Information Use

Table 2 - Avg Raw Numbers Seen		
	Frame	
Evaluating	Negative	Positive
City_male	14.15	85.19
City_female	15.92	85.74
Self_male	14.52	83.96
Self_female	13.62	85.81

Table 3 – Negativity Bias, Gender, and Performance Information Use

Table 3 - Mean Performance Rating		
	Frame	
Evaluating	Negative	Positive
City	70.75	81.96
Self	71.63	82.31

Table 4 – Negativity Bias, Gender, and Performance Information Use

Table 4 - Negativity Bias Differences (Between-group differences highlighted)			
Row Mean-Col Mean	City, Neg	City, Pos	Self, Neg
City, Pos	11.21 0.018*		
Self, Neg	0.88 1.00	-10.33 0.03	
Self, Pos	11.56 0.01	.35 1.00	10.68 0.02*
*Signficiant at p =.05 level			

Table 5 – Negativity Bias, Gender, and Performance Information Use

Table 5 - Mean Performance Rating		
	Frame	
Evaluating	Negative	Positive
City_male	75	80.15
City_female	66.32	84
Self_male	71.12	81.65
Self_female	72.07	82.96

Table 6 – Gender, Negativity Bias, and Performance Information Use

Table 6 - Negativity Bias Differences (Between-group differences highlighted)							
Male				Female			
Row Mean-Col Mean	City, Neg	City, Pos	Self, Neg	Row Mean-Col Mean	City, Neg	City, Pos	Self, Neg
City, Pos	5.15 0.73			City, Pos	17.68 0.01*		
Self, Neg	-3.88 0.87	-9.03 0.28		Self, Neg	5.75 0.73	-11.93 0.16	
Self, Pos	6.65 0.54	1.5 0.99	10.53 0.16	Self, Pos	16.64 0.02	-1.04 1.00	10.89 0.20
*Signficiant at p =.05 level							

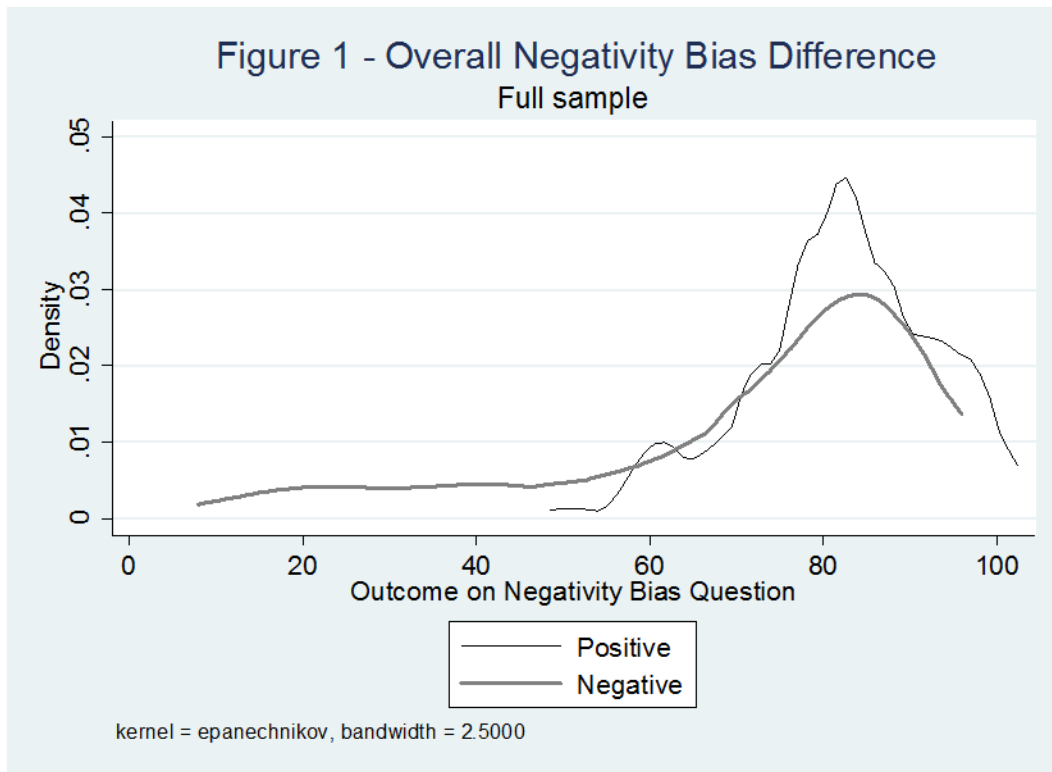Figure 1 – Negativity Bias, Gender, and Performance Information Use

Figure 2 – Negativity Bias, Gender, and Performance Information Use