APPLYING A BEHAVIORAL APPROACH TO THE USE OF PERFORMANCE

INFORMATION BY PUBLIC MANAGERS

Sean Webeck

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the School of Public and Environmental Affairs,
Indiana University
July 2018

ProQuest Number: 10837534

ProQuest 10837534

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____
Sean Nicholson-Crotty, Ph.D.

_____
William T. Bianco, Ph.D.

_____
Jill Nicholson-Crotty, Ph.D.

_____
Asmus Olsen, Ph.D.

_____
Tom Rabovsky, Ph.D.

Date of Oral Examination
June 21, 2018

ii

To my wife, Jennifer. I could not have done this without you.

ACKNOWLEDGEMENTS

During the process of earning a PhD, one makes progress in a number of areas: intellectual, personal, professional, social, etc. In my own experience, this is notably the case for the dissertation phase. As I reflect on the many things I learned in my time at Indiana University, I think the most important school-related thing I learned—the one that has changed the way I approach the world the most—has been to explore the assumptions we use as we go about our lives. I appreciate that this knowledge is something that rings just as true outside of academia as it does in.

GIVING THANKS

In the process of acknowledging growth, one must also show gratitude for those who assisted them in the various areas of their life in which they feel they have grown. For me, I want to thank several people.

Faculty

Early in my doctoral studies, several faculty members helped me find my way. Dr. Sameeksha Desai and the late Dr. Evan Ringquist believed in me when I applied and in the early stages of my PhD program. Drs. Mike McGinnis and William Bianco from the IU Department of Political Science also played an important role in pushing me towards theoretical questions in political science while I was still in the early years of my PhD program. Dr. McGinnis advised me in the process of navigating my way out of what I thought I was interested in towards related topics that allowed for better research. And, I want to thank Dr. William Bianco for what I learned about institutions in his Legislative Politics course as well as his service on my committee.

Dr. Matthew Hayes was one of the most formative figures in my academic development at IU. His Political Psychology course literally changed the research I was interested in and the way I look at the world. The ideas for much of my current research interests, including my dissertation, came out of interacting with him.

Dr. Asmus Olsen provided vital expertise on the topic of my dissertation. I attended a panel he was on at the 2015 Midwest Political Science Association's Annual Conference. He presented research on how citizens exhibit cognitive biases in the use of performance information. I had been thinking about similar questions pertaining to public managers because of my interactions with Matthew Hayes. Asmus responded positively to my ideas when I first shared them with him and he continues to support my research by providing timely feedback that improves the substance of my products.

From SPEA, I want to thank several individuals. Dr. Tom Rabovsky. Not only is Tom on my dissertation committee, he has shown support throughout my dissertation project by pushing me to think more deeply about what I am trying to say and how I will communicate it to others. Dr. Jill Nicholson-Crotty has facilitated my professional development in several ways. One key way has been through our research collaborations over the past few years.

Finally, from SPEA, I want to especially thank Dr. Sean Nicholson-Crotty. He is the first mentor I've had in my life. As my advisor, Sean has been instrumental in my professional development. He pushed me when I needed it. Without his support I doubt that I would have been able to be offered my first position. I also appreciate having had the opportunity to learn *with* him, as the dissertation idea I brought to him was one that required both of us to learn new material.

Family and Friends

Of all the people I want and need to thank, none are more important to me—none have been more supportive and helpful to me through the PhD process—than my wife, Jennifer. I truly feel that my wife and I have learned together through this process. Her selflessness and support have allowed our family to grow and for me to succeed during our six years in Bloomington. All the while, I've been able to be by her side and watch her learn and explore. It is truly humbling to see all that she does.

I am impressed by her strength, energy, and the way she interacts with others. While here in Bloomington she earned a master's degree—something she never thought she could or would do. And, she did it while having our fourth child while maintaining a full-time student status to remain eligible for the nationally competitive fellowship she earned. She has since developed her career. I enjoy seeing her continued success. I know that I should eventually stop being surprised as she continues to reach new heights.

When I think back upon the milestones of graduate school I think about where my children were in their lives during those events. I am proud to have had the opportunity to watch the four children who've spent the past six years in "Apartment 214" grow. I look forward to watching all of them continue to grow in their own unique and individual ways.

Aurora, who started in pre-school during our first year, has grown into a person trying to understand her place in the world. I enjoy being able to watch her learn about new things that excite her. I love seeing her process the world in front of her. She has helped me to learn compromise and patience, she has helped me learn to be a father.

I love watching how much Zoë enjoys nature and the outdoors. I appreciate being able to watch her experience life. Her positivity is uplifting. I will always cherish our trip to San Francisco over the Labor Day weekend of 2017.

Liam, who I love for his sweetness and his energy, will always be dear to me. He has taught me it is OK to be who we are. He is helpful, even when he does not fully comprehend the things that limit my ability to fully interact with him. His love is unconditional.

Trajan is our little ball of energy. I love to wrestle and "play foreheads" with him. I love his questions and the thinking that goes into them, even for a little guy. He helps me remember to stop and enjoy the time we have on this earth.

Six people, four kids, and two graduate degrees take a lot of time and require a lot of attention. Nonetheless, I'd like to think that Jenny and I have been able to stay committed to living a life filled with "fun and adventure". During our time here, we took two trips to the Outer Banks. We had a big family trip out West—we saw the Tetons, Yellowstone, Mount Rushmore, visited Wall Drug, saw the Badlands, and other things. And, Jenny is very good at taking the kids on trips to see various attractions. We also enjoyed some international travel. Right before arriving in Bloomington we (all but Trajan) lived in Jordan and took a trip to Turkey, which fueled a desire in Jenny and me to learn some Turkish. We (all six of us) took a trip to Paris and Amsterdam both for our tenth wedding anniversary and because I was invited to a conference in Amsterdam. And, this past Christmas, we spent two wonderful weeks in Denmark. I truly hope we can continue to be committed to seeing the world and sharing that with our children.

I cannot express the appreciation I have for the love and support these people have shown me despite my physical limitations during our time in Bloomington. I know it is not always easy. And, they have all helped in their own ways during my recovery from hip replacement surgery at

Sean Webeck

APPLYING A BEHAVIORAL PERSPECTIVE TO THE USE OF PERFORMANCE

INFORMATION BY PUBLIC MANAGERS

In the public sector, performance management systems are designed to create and then deliver performance information to key decision makers to inform decision making and improve organizational performance. Despite the growing popularity of these systems, we have a very limited understanding of how public managers actually use the information performance management systems produce (Moynihan, 2015). A key premise of this project is that existing work has relied too little on experimental methods to understand decisions regarding performance information.

The long-term objective of this project is to apply a behavioral perspective to our understanding of how public managers use performance information and more generally to the study of bureaucratic decision making. More specifically, this dissertation looks within the process of decision making to understand what factors might lead to variation in the interpretation of performance information. Using survey experiments among public managers, it tests two competing frameworks for how people make decisions. According to the "rational" school, people make decisions in a way that seeks to maximize their expected utility (Von Neumann and Morgenstern, 1944). On the other hand, the behavioral framework suggests conditions under which people make non-maximizing decisions because of the limits we face as people and as social animals (Simon, 1947; Cyert and March, 1963; and Kahneman, 2011b).

This dissertation tests for evidence of the assumptions of these schools in several experiments: negativity bias, historical and social comparison biases, and gender and cognitive bias. Results of the survey experiments demonstrate systematic variation in the interpretation of

performance information, suggesting evidence in support of using a behavioral perspective in

future research on performance information use by public managers.

_____

Sean Nicholson-Crotty, Ph.D.

_____

William T. Bianco, Ph.D.

_____

Jill Nicholson-Crotty, Ph.D.

_____

Asmus Olsen, Ph.D.

_____

Tom Rabovsky, Ph.D.

TABLE OF CONTENTS

Curriculum Vitae

xi

Performance Information—Facts and Interpretations: A Performance Information Processing Framework

Sean Webeck

Indiana University

Abstract The question of how public managers use public sector performance information has received a lot of scholarly attention in recent years. The promise of performance management systems was to rationalize the decision making process by creating objective performance metrics that citizens, political officials, and public managers could use to assess the performance of public organizations. Some theoretical work suggests, however, that there is a certain subjectivity to these data, which arises from an individual's role in their organization or broader political environment. Furthermore, a recent spate of experimental work in this area suggests subjectivity might also arise through cognitive bias. I bridge these two bodies of scholarship with a framework of performance information processing, which incorporates four models of political information use into the story of how public managers use performance information. I suggest that cognitive bias can contribute to the subjectivity of performance information in the way that public managers *process* performance information. In other words, a model of *meaning avoidance* suggests that managers accurately receive performance information from management systems, but that cognitive biases influence the ways in which they interpret or act upon that information. In this essay, I demonstrate the evidence for the first three steps of this model. Specifically, in three separate experiments I show that despite different presentations, public managers can accurately recount the objective information they saw when asked to recall it. I also provide evidence (more fully demonstrated in Essay 2 and Essay 3 of this dissertation) that despite being equally aware of objective raw performance metrics that they exhibit evidence of cognitive bias when asked to interpret the meaning of that information. This study contributes to the broader discussion of how individuals use performance information.

Introduction

"*Empiricism assumes that objects can be understood independently of observing subjects. Truth is therefore assumed to lie in a world external to the observer whose job is to record and faithfully reflect the attributes of objects.*" Harvey (2001)

"*Against positivism, which halts at phenomena—"there are only facts"–I would say: no, facts is precisely what there are not, only interpretations.*" Friedrich Nietzsche, as cited in Cox (1999)

How public managers use the performance information created by now ubiquitous performance management systems is one of the big questions facing public management scholars today. One of the challenges facing scholars interested in tackling this question is our ability to adequately model the broader set of actions and processes that ultimately contribute to performance information use. In 1995, Robert Behn suggested that one of the big questions facing scholars who study public sector organizations was understanding how "*public managers use measures of the achievements of public agencies to produce even greater achievements*" (Behn 1995). Yet, recently, 20 years after Behn pointed out a major question for researchers in the field, a prominent scholar commented that "we know little about the basic tendency of individuals to incorporate and use performance information" (Moynihan 2015). This is not to say that scholars have not developed frameworks, such as the Interactive Dialogue Model (IDM), that significantly contribute to our understanding of performance information (Moynihan 2008). However, there is more to learn, as evidenced by the fact that, in the decade since the IDM's initial publication, we have developed more key insights into the phenomenon of interest, but as a community we have yet to adequately update our dominant models.

More recently, a new research program on the behavioral foundations of performance information use has contributed a rash of empirical evidence for how individuals process performance information. This behavioral turn has yielded important insights because, typically,

2

scholarship on how public managers use performance information had relied on survey responses and self-reported information (Ammons and Rivenbark 2008, Kroll 2015, Moynihan et al. 2017).[1] Behavioral research often uses different methodologies and theoretical foundations from research based in organizational theory, allowing for different insights. Following this behavioral approach, I aim to contribute to the understanding of how public managers use performance information in two ways. First, I argue the information processing approach allows us to gain new leverage on the question of performance information use because it requires us to look at individual steps in the way that an individual *processes*, or make decisions about, information (Oppenheimer and Kelso 2015). Second, I suggest the importance of the role of the *interpretation* of performance information as a cognitive step that creates subjectivity in the use of objective information generated by performance management systems (Gaines et al. 2007).

Insights into the psychological factors that influence how public managers process performance information may have implications for performance management. On the one hand, how public managers *interpret* performance information potentially influences how public managers "use" performance information, but current scholarship has not paid adequate attention to this important antecedent. On the other hand, an information processing approach, rooted in psychology, also suggests important limitations in dominant models of performance information use, such as the Interactive Dialogue Model (IDM). Specifically, while current models allow for certain types of subjectivity, say those arising from organizational factors, they cannot adequately incorporate well established cognitive biases in human decision making.

I borrow a framework from political psychology on political information use to address this shortcoming (Gaines et al. 2007). The Gaines et al. framework of political information use

---

[1] Typical approaches include single-city case studies, multicity surveys, and multicity case studies.

consists of four cognitive processing models that seeks to describe how individuals process information over multiple steps. Here, I offer a framework of how individuals process *performance information*, offering "awareness" and "interpretation" as distinct steps in how people process this type of information. This approach gives us both more descriptive and predictive power towards developing theory in this area.

As I will show in this essay, while individuals tend to be equally aware of the objective value of the performance information they observe, their interpretations are prone to cognitive biases that arise in large part from the way in which information is presented to them. That is, subjectivity arises specifically from the *action* and *process* of interpretation and not directly from the information itself nor in how individuals initially incorporate new information. Taken together, this model and relevant empirical findings allow for a better understanding of how the design of performance measurement systems can influence performance information use. And, since empirical findings from the broader study of performance information use suggest there is value in updating the IDM, this framework can help students of public management begin to synthesize prevailing models with evidence from how individuals—citizens, politicians, or public managers—use performance information.

In this essay I empirically test the claim that individuals are objectively aware of the performance information they observe. I also offer a brief description of the empirical results of one test for interpretation to demonstrate the validity of the larger claim that it is the process of interpreting performance information, and not how individuals receive performance information, that is subject to cognitive bias. Essay 2 and Essay 3 of this dissertation provide more conclusive evidence that bias arises during the interpretation process.

4

In what follows, I review early work on performance management, emphasizing the question of performance information use. I then review and provide a discussion of Moynihan's (2008) Interactive Dialogue Model, including its key assumptions. In noting some shortcomings of the model, I suggest how an understanding of human behavior, as well as recent experimental evidence on the study of performance information use, provide support for incorporating an information processing perspective into the IDM. I point to work from Gaines and colleagues (2007) to potentially address these challenges and refine the larger IDM. I follow with a set of general propositions and specific hypotheses for ways in which cognitive biases can influence how individuals process performance information. Here (and in the other essays in this dissertation), I then provide empirical evidence for this framework. I conclude with a discussion of these results and by highlighting avenues of future research.

Literature Review
Performance Management Systems

When it comes to the public sector, performance management regimes are everywhere. As one scholar noted, "the dissemination of quantitative measures of performance has been one of the most widespread trends in government in past decades" (Moynihan, 2015, 33). And, throughout the performance management movement, leading scholars have tried to direct attention to the question of how public managers use performance information (Behn 1995, Moynihan and Pandey 2010). Yet, recently, a prominent scholar commented that "we know little about the basic tendency of individuals to incorporate and use performance information" (Moynihan 2015). If we are to accept the claim that how bureaucrats use performance information is an important question for public management researchers (Moynihan and Pandey

5

2010), we must acknowledge some of the limitations of previous research to theorize and teach future practitioners about the subject (Kroll 2015, Moynihan et al. 2017).

It is important to acknowledge there are three primary groups of users of performance information: citizens, political officials, and public managers (Van Dooren and Van de Walle 2011). Then, within those studies that focus on the question of how public managers use performance information, one group of studies focuses on the question of what it means for public managers to use performance information (Behn 2003, Moynihan 2010, Van de Walle and Van Dooren 2011, Van Dooren and Van de Walle 2011). Another looks at how public managers respond to performance information—that is, *how* do they use it? A recent review of 25 empirical studies found several organizational factors which regularly contributed to performance information use: measurement system maturity, stakeholder involvement, leadership support, support capacity, innovative culture, and goal clarity. The review also highlighted areas that needed more research attention, such as developing our understanding of performance information use through methodological and theoretical work. See Kroll (2015) for a complete review of this work. Finally, there have been attempts to develop theoretical frameworks for how public managers use performance information (Moynihan 2008, Meier et al. 2015). In this essay, I contribute to work that fits within these latter two groups of studies.

Performance Management Systems as Decision Making Systems

Moynihan defines performance management as "a system that generates performance information through strategic planning and performance measurement routines and that connects this information to decision venues, where, ideally, the information influences a range of possible decisions" (2008). In this view of performance management systems, to understand how public managers use performance information, we must understand how they *make*

6

decisions, about performance (information). There is growing evidence of the value of incorporating an information processing approach when considering the subject of decision making. This general approach suggests *basic models of cognition should form the basis for how we conceptualize human decision making*. These models allow us to focus our attention on "how decision-relevant information is sampled, retrieved, and integrated" (Oppenheimer and Kelso 2015, 283).

An Interactive Dialogue Model

In 2008, Moynihan laid out what he referred to as the "Interactive Dialogue Model" (IDM). This model describes how and why public managers use performance information and has become perhaps the dominant framework that scholars employ when asking these questions. While scholars have presented different takes on what it means to "use" performance information—see, for example, Behn (1995)—Moynihan (2008) suggests the purpose of use is ultimately to persuade. In line with the idea of persuasion as the aim of use, at the time of its publication, the most important takeaway from the model was that performance information can be subjective. Specifically, in the IDM, performance information is ambiguous because of political considerations *a priori* to any descriptive story of performance information "use". This idea of subjectivity is contrary to the performance management doctrine, representing a major break in the theoretical development of performance information use by public managers (Moynihan 2008).

Elements of the Model

There are three fundamental elements of the model. They are: (1) performance information, (2) the individual decision maker (i.e., public managers), and (3) the environment(s)

7

in which these other elements exist or operate. Here, environment is meant to imply both the organizations in which individuals work and the political environment(s) in which those individuals and organizations are situated. The model's emphasis on organizational and environmental factors (see Moynihan 2008, p. 103) parallels many other studies in the research program (Kroll 2015).

Taken together, these three elements have shaped the way many scholars, including Moynihan, have looked at performance information use. Grounded in the logic of institutions, the IDM gives us a story in which institutions matter. Yet, the emphasis on institutions raises the question of what other elements might influence and help explain how public managers use performance information. One potential avenue of explanation is the relationship between performance information use and factors at the level of the individual (Kroll 2015).

Assumptions of the model
There are six basic assumptions to the model (Moynihan 2008, 102). First, performance information is not comprehensive. Second, performance information is ambiguous. Third, performance information is subjective. Fourth, the production of performance information does not guarantee use. Fifth, institutional affiliation and individual beliefs will affect selection, perception, and presentation of performance information. Sixth, the concept of dialogue will affect the ability to use performance information to develop solutions.

In this model, organizational and political factors play a significant role in shaping performance information use. They influence performance information (use) in several ways. This includes, the presentation of performance information, whether an individual considers (i.e.,

"looks at") performance information, how they interpret performance information, and, finally, how they "use" performance information.

To be clear, the IDM assumes performance information is subjective because: (1) individuals can choose to present information subjectively, (2), even the act of considering performance information is, in and of itself , a choice, (3), individuals interpret performance information based upon organizational and political factors, and (4) individuals use performance information to strategically achieve organizational and personal objectives. These assumptions lead to a meta-assumption: "that simply because performance information exists, there is no guarantee that it is used" (Moynihan, 2008, 102). I want to push on this meta-assumption because I think it oversimplifies how human beings process information in two key ways. First, I think it treats use as binary—either people use it, or they do not.

Second, here, use is implied to be a discrete act. This idea of use is in line with early writings from Behn (1995). In this sense, use is an action taken whereupon said action has been informed by the performance information in question (Van Dooren and Van de Walle 2011). In other words, to be adequately considered as performance information "use", a decision maker must have looked at the performance data in question and both 1) become aware of the performance information and 2) updated (or not)[2] their interpretation of a policy area. The key assumption here—and a shortcoming of the IDM in its current form—is that this process is not well explained or described. Rather, it assumes that once an individual decides to look at performance information that she and all others will incorporate and interpret that information in a uniform manner. On one hand, characterizing use in this way facilitates observation. On the other, it does not seem to hold up to some basic assumptions from information processing theory

---

[2] This part could be flexible in a theoretical sense, but it is left unspecified in the IDM.

9

(Oppenheimer and Kelso 2015). To clarify this last point, the existing work makes critical—yet unstated—assumptions about how individuals process performance information.

Psychology and the Subjectivity of Performance Information

Another reason performance information might be subjective can be found in the way that human beings process information. While Moynihan points to a potential role for psychology in a confirmation bias (steps 2 and 4 in Figure 1), the IDM is largely bereft of psychology as an influence in the larger process of using performance information. The closest Moynihan gets to this argument is when he makes this falsifiable hypothesis: "Different actors can examine the same performance information and come up with competing, though reasonable, arguments for what the information means" (2008, 113). Carrying on with this line of thinking, Moynihan says:

> "Performance information does not necessarily result in clearer decisions if the actors involved cannot agree on what it tells them about current performance, changing budgets, or management. As roles motivate the actors involved to understand performance information differently, the inherent ambiguity in performance information will be exploited." (2008, 16-17).

In terms of predictive power, the model stops there.

But, the preceding statement makes three key assumptions which should be explored in more detail. Because ambiguity arises through roles there is little space for other factors to contribute to subjectivity. Yet, recent empirical work suggests cognitive processes may also contribute to performance information's ambiguity and subjectivity. The standard take on bounded rationality is that human beings have significant constraints on their ability to hold and process information. In the IDM, however, information overload leads individuals not to "try to process all information but select information that they find useful" rather than simply being

10

unable to accurately process information they have (2008, 17). Thus, the first key assumption here is that selection is a deliberate (cognitive) act. Second, there is a question about whether actors can even agree on what performance information tells them. A third and related assumption is that the ambiguity inherent in performance information arises because of a deliberate act on the part of the individual. Public management scholars should give more attention to these assumptions.

If we can learn from work on information processing and heuristics, there are several reasons why individuals respond differently to the same information (Gigerenzer and Gaissmaier 2011, Oppenheimer and Kelso 2015). And, many of these different interpretations arise not from deliberate cognitive processing but from a type of cognitive processing that relies on speedy, snap decisions

Behavioral Foundations of Performance Information Use: Empirical Evidence
Recent empirical work in public management suggests the value of considering the behavioral foundations of performance information use. Some scholars have pointed out that behavioral factors might produce systematic variation in the use of performance information among bureaucrats (Kroll 2015, Moynihan et al. 2017). These, and other studies, suggest psychology may play a role in helping us understand how public managers use performance information (Moynihan 2008, Salge 2011, Nielsen 2013, Kroll 2015, Moynihan 2015, Andersen and Moynihan 2016). In addition to these, some very recent pieces also demonstrate the utility of incorporating an individual-level behavioral approach to examine the use of performance information. Other studies suggest this perspective can contribute to our understanding of how individuals, broadly considered, respond to performance information. These include studies on

citizens (Olsen 2013, Andersen and Hjortskov 2015, Olsen 2015, Baekgaard and Serritzlew 2016, Barrows et al. 2016, Hvidman and Andersen 2016, Olsen 2017) and politicians (Olsen 2014, Nielsen and Baekgaard 2015, George et al. 2016, Nielsen and Moynihan 2016). This approach should be understood within a broader context of seeking to understand the psychological foundations of public administration (Grimmelikhuijsen et al. 2016).

For our purposes, there are two important takeaways from these studies. First, experimental methods are a useful approach to develop our understanding of performance information use across a variety of political actors (Anderson and Edwards 2015, Bouwman and Grimmelikhuijsen 2016, Jilke et al. 2016, James et al. 2017, Moynihan et al. 2017). Second, when it comes to performance metrics, these studies suggest that, depending upon the circumstances, individuals exhibit various cognitive biases and utilize several heuristics when responding to performance information. Evidence for cognitive bias in the use of performance information by various actors supports the value of taking an information processing approach (Gigerenzer and Gaissmaier 2011, Oppenheimer and Kelso 2015).

One cognitive bias that has received some attention from public management scholars in recent years is motivated reasoning. It is now well established that "motivation may affect reasoning through reliance on a biased set of cognitive processes: strategies for accessing, constructing, and evaluating beliefs" (Kunda 1990). Epley and Gilovich argue that our motivations potentially influence the way we process information in one of two ways (2016). First, our biases and preferences might lead us to avoid certain information or to emphasize other pieces of information. Second, once we have information, we are free to interpret how we like.

Public management scholars have only recently turned to motivated reasoning to try and understand how individuals use performance information. One study shows how political

12

motivations influence how political officials prioritize goals in the face of the information presented to them (Christensen et al. 2018). Another set of studies suggest that individuals' motivations influence the way they interpret performance information (Baekgaard and Serritzlew 2016, James and Van Ryzin 2016, Baekgaard et al. 2017). Unfortunately, none of these studies adequately disentangle the question which arises from Epley and Gilovich—namely, do differences in interpretation exist because of deviations in *acquiring* or *interpreting* performance information? While this question arises specifically in the context of a discussion about motivated reasoning, scholars could apply a wide array of cognitive biases to the question of if and how individuals acquire and interpret performance information.

Theory

The value in thinking about how public managers process performance information is that it allows us to think about systematic variations in the cognitive process of information use. One approach from political psychology that may help to inform our understanding of how public managers use performance information considers how individuals move from facts about politics to political opinions (Gaines et al. 2007). Whether individuals are aware of political facts and how they form political opinions are fundamental questions in political science. Similarly, scholars of public management are interested in how well public managers understand the performance of their organization. And, accountability problems lead public management scholars to focus on how managers make decisions. Ergo, citizens, political officials, public managers and scholars are all interested in the extent to which there is a connection between facts about public sector organizations and decisions public managers make relating to that information.

13

Four Models of Political Information Use

Gaines and colleagues (2007) suggest four models of processing political information in which a) objective information exists in the larger political environment, b) individuals become aware of the information, c) individuals must interpret the political information, and d) finally, individuals must arrive at policy positions. Their framework builds off work that seeks to understand "whether people update and what it means to update" (*ibid*., 958). Some work argued that citizens were able to objectively update their policy views when new political information entered the environment (Gerber and Green 1998, Gerber and Green 1999) while other work showed evidence of bias in policy views after new information became publicly available (Bartels 2002, Taber and Lodge 2006).

According to Gaines and colleagues (2007), the interpretation of political information plays an important role in moving from political fact to political opinion because interpretation represents the step in the cognitive process where individuals give meaning to political information. In more colloquial terms, partisanship is a strong drug that significantly influences the way that individuals look at—i.e., interpret—political information (Kunda 1990, Taber and Lodge 2006). But, they did not advance this same expectation about the ability of individuals to acquire new political information and be aware of it. That is, they did not expect that partisanship influences an individual's ability to accurately update their understanding of political facts.

More recently, the question of information acquisition has interested scholars because of the discussions of "alternative facts" and "fake news" in the broader political environment. And, as suggested by Epley and Gilovich (2016), there is some evidence that motivated reasoning does in fact influence how individuals receive and update on *political* facts (Nyhan and Reifler

14

2010, Hochschild and Einstein 2015, Yeo et al. 2015, Schaffner and Roche 2016). Without a doubt, the role of bias is an important question in the acquisition and awareness of political information. And, scholars of public management should also inquire about the factors that lead to deviations in the acquisition process when considering performance information. Yet, when Gaines et al. (2007) demonstrate empirically that cognitive bias influences how individuals process political information between steps "b" and "c", they provide evidence that cognitive biases can arise from the process of interpretation.

I expect awareness after information acquisition to be considerably less susceptible to the influence of cognitive bias for performance information than for political information. This expectation arises because, at least ostensibly, performance information is objective in a way that political information inherently is not; despite the objections about the objectivity of performance information, the original intention of performance management systems was to create measures of performance that accurately and adequately captured the function in question. For example, telling someone 75% of students passed a Math exam in a school or that 86% of residents expressed satisfaction with a city's road maintenance efforts is likely to invoke less partisan or ideological filtering than telling an individual that 97% of scientists agree with the anthropogenic causes of climate change. As such, I expect the model of performance information processing which emphasizes how cognitive bias can influence the interpretation, rather than acquisition of, performance information (Model 3, discussed below), represents the best model for understanding how individuals (e.g., public managers) process performance information. In the following section, I modify their framework to apply to performance information processing.

A Framework of Performance Information Processing

If we consider the similarities between political information and performance information—namely, they are information—then the Gaines framework could apply to performance information use (by citizens, political officials, *and* public managers) as well. Ergo, how public managers interpret performance information might have an important effect on the decisions they make.

As it stands, interpretation plays a role in the IDM when managers choose to consider information or how they might spin performance information to the benefit of their organization. But, the IDM does not consider how individuals interpret performance information in line with the experimental evidence which suggests the role of cognitive bias in the use of performance information. It does not consider the role of information processing in how managers *interpret* performance information. Let's suppose a manger wants to consider and then use performance information. What might this process look like?

Table 1 shows an adapted version of the four models of information updating proposed by Gaines and colleagues (2007)—I suggest referring to them individually as performance information processing models or the Performance Information Processing Framework (PIPF). The models have been adapted to portray four steps of the updating process for performance information use. First, I discuss the different steps and then I discuss the four models which describe different processes of performance information use.

The first step represents the raw performance metric; a manager sees performance information. In the second step the manager becomes aware of the metric; they can recall the information they just saw. In the third step, a manager interprets the information; the manager gives meaning to the performance information and updates their belief about an organization's

performance.   Finally, the manager decides to undertake some action based upon this information; in the lexicon of our field, the manager has "used" the information.

Table 1 - Four Models of Performance Information Use								
Model	Model Name	Step 1	Process	Step 2	Process	Step 3	Process	Step 4
1	*Complete updating*	Performance Information	→	Awareness	→	Interpretation	→	"Use"
2	*Fact avoidance*	Performance Information	‡	Awareness	→	Interpretation	→	"Use"
3	*Meaning avoidance*	Performance Information	→	Awareness	‡	Interpretation	→	"Use"
4	*"Use" disconnect*	Performance Information	→	Awareness	→	Interpretation	‡	"Use"
Here are four models of performance information use along with relevant steps and representations of cognitive processing in the model. The "→" sign signifies a smooth transition from one step to another. The "‡" sign represents a cognitive process that does not cleanly flow from one step to the other. For example, in Model 2, individuals may see performance information but not become aware of it (i.e., they would not be able to recall what that information was).								

In Model 1, *complete updating*, there is a smooth transition from performance information, to awareness, to interpretation, to use. That is, if a manager sees new information which is contrary to her original understanding of the performance of the organization (Meier et al. 2015), she becomes aware of it and interprets that information in a way that accurately reflects that information. In this case, we would expect her to recall and then provide an interpretation that matches or is at least very similar to the performance information she saw. The information is then used in a way that reflects the smooth cognitive transitions from step one to steps two and then three.

Model 1, *complete updating*, reflects early thinking on performance information use, which Moynihan refers to as the performance management doctrine (2008). In other words, for many people this would probably be the normatively preferred model of performance information use. But, previously discussed empirical evidence from experimental work suggests some form of deviation from this ideal—in all three groups of users. So, we must look to some other descriptive model to help us understand how public managers process performance

information. A model of cognitive processing allows us to pinpoint where, both in a descriptive and a causal sense, the sources of where these variations arise.

In Model 2, *fact avoidance*, managers see performance information but do not update their awareness of what that information was. Gaines and colleagues say some conditions which might lead to fact avoidance are "willful or accidental ignorance" or "if changing conditions create mental discomfort" (2007, 960). These circumstances might lead people to simply pay less attention to reports of performance changes. In the case of performance management, this might be when a public manager is deeply invested in a project that is being evaluated negatively or if they get information that is drastically different than their worldview. The examples I investigate in this project should not lead to those kinds of cognitive conditions or challenges. This could be an area of interest to scholars in the future.

In Model 3, *meaning avoidance*, managers see performance information and become aware of it but do not change their interpretation of the information. That is, the way they interpret the information does not flow smoothly from the newly acquired information itself. Gaines et al. (2007) illustrated this model by showing how partisanship influenced how Americans interpreted the Iraq War. Even though co-partisans representing the two predominant U.S. political parties were aware of the increase in troop casualties, Republican co-partisans (with a Republican as the sitting-President) interpreted this to be less severe than Democratic co-partisans. They suggest that motivated reasoning or reference biases might represent cognitive biases that lead to meaning avoidance, but I argue there are a wide number of cognitive biases that might affect the way that public managers interpret performance information.

In Model 4, *"Use" disconnect*, managers would become aware and then interpret the performance information but would ultimately not use it in a way that flows smoothly from the

18

interpretation stage.   This could happen for a variety of reasons—some individual, others institutional.   Ostensibly, the IDM captures this approach.

Expectations Arising from the Framework

Using the logic of the Performance Information Processing Framework as well as recent empirical findings on the behavioral foundations of performance information use, I put forward the following general hypothesis: when faced with performance information, public managers will process that information in a way that deviates from the "complete updating" performance information processing model.   In the past, systematic deviations from rationality in the assessment of performance information have generally been "the bar" of evidence necessary for claiming that cognitive bias influences the way that individual's use of performance information. Now, the PIPF allows us to move beyond this limitation and to specify *where* in the cognitive process these biases arise.

Evidence of the role of cognitive bias in the processing of performance information could come in three forms.   First, individuals might not adequately update their awareness of the objective information (i.e., *fact avoidance*).   Second, if they hold an accurate awareness of what the performance information is or says, they may simply interpret it in a way that diverges from a smooth transition from step 2 to step 3 (i.e., *meaning avoidance*).   Third, if there are no deviations from the original information after individuals interpret it, other factors could lead individuals to use it in a way that does not align with the considered performance information (i.e., "use disconnect").   As highlighted in the IDM, role, or other organizational factors, could be one example of a factor that might lead to *use disconnect*.   And, while future research should

investigate this step in the Performance Information Processing Framework, for the remainder of this essay the emphasis is on Model 2 and Model 3.

The Gaines et al. framework suggests that Model 3, *meaning avoidance*, will have the best descriptive power of these models. For this reason, I expect that when given performance information, public managers will be able to accurately recall—i.e., they can demonstrate an "awareness" of—previously observed performance information. But, these same individuals, faced with the same conditions and performance information of which they are aware, will be prone to exhibit cognitive biases when asked to interpret this information. I pre-registered these expectations with the Evidence in Governance and Politics (EGAP) group, application ID: 20180425AD.

Beyond this general expectation, I am going to provide three specific hypotheses based upon cognitive biases. These primarily rely on reference points, which are "stimuli of known attributes that act as standards against which other categorically similar stimuli of unknown attributes are compared in order to gain information" (Yockey and Kruml 2009, 97). The biases which I will attempt to elicit arise out of recent studies in this area (Charbonneau and Van Ryzin 2015, Meier et al. 2015, Olsen 2015, 2017). I chose these because I wanted the first tests of the PIPF to study biases that would potentially be of interest to a broad group of public management scholars, such as those that rely on comparisons. In this essay, I will test step 2 of the PIPF (i.e., Model 2). Specifically, I will look to see the extent to which public managers can accurately recall objective performance information (even after being asked to provide an interpretation of this information).

H1: When faced with performance information in the context of performance benchmarks and justification requirements, public managers will be able to accurately recall the objective performance information.

H2: When faced with performance information and a historical performance comparison, public managers will be able to accurately recall the objective performance information.

H3: When faced with performance information and a social performance comparison, public managers will be able to accurately recall the objective performance information.

The reader will notice that I am only testing the hypothesis for step 2. There are two reasons for this. First, I need to explicitly test the awareness step. Second, I've already tested the interpretation step. Those results can be found in Essay 2 and Essay 3 of this dissertation. Nonetheless, I will provide a brief empirical overview of one of the interpretation tests here as well to demonstrate the value of performance information processing Model 3, *meaning avoidance*.

Data and Empirics

Empirical evidence to test these hypotheses come from three experiments run over two surveys. In the first experiment, I tested how individuals with significant public sector work experience process performance information in the form of business satisfaction rates in a city (also see Essay 2). I delivered experiments 2 and 3 in a single survey that asked respondents to make a performance assessment given comparative (historical and social) performance information (also see Essay 3). For each of these three experiments, I ask individuals to interpret performance information and then, at a later point in the survey, respond to a question asking them to name the raw performance metric they saw. In this way, I can manipulate the temporal

stages of the Performance Information Processing Framework to test the validity of the *meaning avoidance* model.

Experiment 1

Data

I used surveys to collect data for this experiment in three phases. Individuals were paid for their participation in all phases. I designed data collection instruments for each phase using Qualtrics. I 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 (Litman et al. 2016).

In the first phase, I ran a short survey that allowed me 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.). I provided representative examples in case the sector type would confuse anyone. In addition to this question, I also asked individuals if they had ever worked in each of the three sectors. Respondents could select "yes" or "no" to specific (individual) questions about each sector. 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. 5342 unique individuals completed this 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

22

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. 1202 individuals met these qualifications.

I then sent a second survey to these 1202 individuals. This survey included demographic items and scales for the Big 5 personality items and public service motivation. Yet, the real motivation behind this phase was to try and screen out those who passed the first phase of the survey but were not in our population of interest—people with significant public sector work experience; especially, public managers. Someone could easily provide inconsistent answers over time. Individuals might lie in one of the two phases because they believe they know what the researchers are looking for. Or, because multiple individuals use the same MTurk account. I undertook this effort in the hope that I could make a stronger claim about our respondents. I received 773 responses from this wave. Of these, 479 met our qualifications through both waves of the survey. These were the potential respondents who were notified of the opportunity to undertake this survey experiment.

Experimental Design
The data from this experiment come from a study which I pre-registered with Evidence in Governance and Politics (EGAP) under the following ID: 20180425AD.

Respondents were asked to provide a response to a vignette about business satisfaction rates. In this experiment, I was interested in the role of performance benchmarks and justifications as potential moderators for how individuals process performance information. Table 2 shows the group assignments across these two treatment conditions. Ultimately, the experiment had a 2x2 factorial design with individuals randomly assigned to one of four groups:

23

(Group 1) control, (Group 2) justification, (Group 3) benchmark, and (Group 4) benchmark and justification.

Table 2 - Experiment 1			
Group Assignments by Treatment			
Treatments		Benchmark	
		No	Yes
Justify	No	1	3
	Yes	2	4

All individuals saw this prompt:

> For this question, imagine that you are the manager of a business development office for a major metropolitan area. Your city just released its yearly performance metrics and, based on this information, the mayor wants to know how you think the city performed over the course of the last year. For some time, business owners were asked whether they were generally satisfied or generally dissatisfied with the city as a place to do business. In 2011, 57% of business owners indicated they were satisfied with the city as a place to do business. The mayor has tasked your office with improving the business climate in the city.

Individuals in the benchmark groups also saw this sentence at the end of the second paragraph of the prompt:

> The goal has been to increase the percentage of business owners satisfied with doing business in the city to 67%.

All individuals saw a randomly generated performance rating which indicated that between 62% and 72% of business owners were satisfied with the city as a place of doing business. Randomizing the observed performance metric is an important component of our test. It allows us to gain a better understanding of how individuals process

performance information across a range of potential performance metrics. This also allows us to use a stationary benchmark.

All individuals were asked to assess the performance of the city as a place of doing business for the past year (based on this data). All individuals had an equal probability of seeing a value that was a) less than the benchmark (5/11), b) equal to the benchmark (1/11), or c) greater than the benchmark (5/11). (Note, the first two groups do not see the benchmark.) Half of the respondents also need to justify their performance assessment. All respondents who will justify their responses are told they will have to perform this task before they see the raw performance metric.

Then, at a later point in the experiment, individuals were asked to provide the raw performance metric they observed in this experiment. This response became the dependent variable of interest in this experiment.

Results[3]

Table 3 provides the means and standard deviation for the value respondents provided for the performance metric they observed. Note, the mean for all respondents was 62.92 but the randomized metric individuals saw had a range of 62-72. I also provide an assessment of the percentage of individuals within each group who listed the exact performance metric they saw. I conducted a one-way ANOVA to determine if the stated value of the observed performance metric (randomized between 62-72) was different across four (a control and three treatment)

---

[3] Here, I show results for the test of "awareness" and then a very brief discussion of the interpretation component of this experiment. A more thorough discussion of the empirical findings for these experiments regarding the interpretation component of the framework can be found in Essay 2 of this dissertation. Of note, the findings on the interpretation part of this experiment indicated the presence of cognitive bias in the use of performance information but the awareness check did not.

groups. There was not a statistically significant difference in the reported value of the observed performance metric across these groups as determined by one-way ANOVA $(F(3,350) = 0.56$, $p = 0.64)$. This test passed Bartlett's test for equal variances $\chi^2(3) = 6.2108$, $p = 0.102$.

Table 3 - Awareness of Observed Satisfaction Rate (Check)				
Treatment Group	Mean	Std. Dev.	N	Pct. with Correct Response
1	62.66	10.13	88	29.55%
2	63.59	12.06	88	43.18%
3	61.86	11.87	92	31.52%
4	63.71	9.71	86	29.07%
Total	62.94	10.99	354	33.33%

In addition to the one-way ANOVA, I also ran an OLS regression that included the dependent variable (value provided), the original raw performance metric, the individual's treatment group, and the outcome score (i.e., the "interpretation" value). This was meant as a harder test. Of note, only the observed satisfaction variable (i.e., the raw performance metric) was statistically significant in the regression model. These results are in Table 4. Taken together, these two tests suggest there is no difference in the level of awareness that individuals had based upon randomization. So, any differences between the groups in their interpretations must come from that specific step in the PIPF.

Table 4 - OLS Regression	
Variable	Business Satisfaction
Observed Satisfaction	0.74***
	(3.90)
Treatment Group	-0.13
	(-0.25)
Outcome Score	0.04
	(0.90)
Constant	10.54
	(0.88)
N	354
Prob > F	0.0001
R-squared	0.0584
Dependent variable is the response to a solicitation for the observed performance metric. *** $p < 0.01$	

Beyond these base checks, I wanted to run a more robust check by comparing all the groups against one another. Those results can be found in Table 5 where I looked at both one-way ANOVA and standard deviation tests for the awareness check across the groups. None of the one-way ANOVA tests showed statistically distinct responses between the groups. Although, one comparison was statistically significant at the $p < 0.05$ level and another at the $p < 0.1$ level in the standard deviation tests. These results appear to be driven by respondents in Group 4 which saw the performance benchmark and the justification requirement. This group had the highest mean but the lowest standard deviation.

27

Table 5									
Reported Raw Observed Raw Performance Metric (Check): Mean and Variance									
By Treatment Class									
Variable	Group	Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
Benchmark	No	63.13	11.11	0.84	0.10	0.7505	1.04	0.7898	176
	Yes	62.75	10.89	0.82					178
Justify	No	62.25	11.03	0.82	1.44	0.2315	1.02	0.9012	180
	Yes	63.65	10.93	0.83					174
By Treatment Group (focus Benchmark)									
Group		Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
1		62.66	10.13	1.08	0.24	0.6279	0.73	0.1365	88
3		61.86	11.87	1.24					92
2		63.59	12.06	1.29	0.01	0.9433	1.54	0.0464	88
4		63.71	9.71	1.05					86
By Treatment Group (focus Justify)									
Group		Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
1		62.66	10.13	1.08	0.31	0.5795	0.71	0.1060	88
2		63.59	12.06	1.29					88
3		61.86	11.87	1.24	1.29	0.2585	1.50	0.0617	92
4		63.71	9.71	1.05					86
Control vs Both Treatments									
Group		Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
1		62.66	10.13	1.08	0.49	0.4862	1.09	0.6967	88
4		63.71	9.71	1.05					86

Experiment 1: Interpretation

Here, I also present evidence on how individuals in this experiment interpreted the performance information they saw. A one-way ANOVA on the assessed organizational performance (i.e., interpretation) showed evidence for a statistically significant relationship for the randomized group assignment ($F(3(350) = 12.31$, $p = 0.0000$). Since we know respondents did not show any statistically distinct patterns in their ability to recall performance information after they were asked to interpret it, these results allow us to confidently say that the randomized group assignment influenced how individuals assessed performance through their interpretations

of qualitatively similar performance metrics but not in their ability to acquire and be aware of this same information. This is significant evidence in support of Model 3, the *meaning avoidance* performance information model.

Experiment 2 and Experiment 3 – Historical and Social Comparisons

Data Collection

Data for these two experiments come from a survey collected from a Qualtrics panel during May of 2017. I was a member of a research team that recruited respondents directly through Qualtrics to avoid some of the potential pitfalls of using other online survey platforms (Stritch et al., 2017). Our research team pre-registered the survey with the Evidence in Governance and Politics (EGAP) group under the following ID: 20170501AC.

Qualtrics screened and provided the respondents for the survey. We provided a stipulation that respondents were professional managers in their organization. The total sample size is 300, with 150 coming from the private-sector and 150 from the public-sector. Our sample includes managers from both sectors because another experiment in the survey required this sector breakdown. All respondents were initially targeted by a partner of Qualtrics through self-reporting. Then, those responses were screened out to remove misidentified respondents using things like red-herrings to make sure the sample is accurate. Qualtrics collected our final data through a partner firm with the ability to prescreen—respondents were asked additional questions at the beginning of the survey to remove individuals whose responses did not match previous identifying responses—and make sure only managers and above could complete the survey.

Experimental Design

www.manaraa.com

Respondents saw two experimental vignettes about performance in the area of education policy. In two separate, but related experiments respondents observed the percentage of individuals who passed standardized tests for English and Math. At the end of the survey, they were asked to provide the raw value of the performance metric they saw. Unlike Experiment 1, which had a randomized performance metric, the raw values in these experiments were fixed.

I modeled the performance information on real test score data to make the experiment more plausible and generalizable to real-world decision making. Specifically, I utilized publicly available data from public schools in the state of Indiana.[4] Data for 2011 and 2012 suggested an average change in the pass rate for English and Math exams to be roughly 2 percent. To determine the raw performance metric individuals would see, I averaged the pass rates for schools in the state for both English and Math standardized exams. Since the experiments also included historical comparisons, I also compared the average change in school test results across time to get a sense of a plausible annual rate of change. Doing this led us to the final performance metrics—77% passed the English exam and 79% passed the Math exam—as well as the historical comparison data (2% change from last year [respondents would see this as a decrease or an increase]).

Due to certain constraints of the research project—namely, recruiting a sample of 300 professional managers—I wanted to address two concerns. The first concern involved the potential for the first experiment to influence the way respondents approached the second experiment. I addressed this by separating the two experiments. Specifically, in this regard, respondents saw *Experiment 2* near the beginning of the survey and *Experiment 3* near the end of the survey. The average response time for the survey across all 300 respondents was 24 minutes

---

[4] https://www.doe.in.gov/assessment/istep-results

and 22 seconds. This meant there was a significant amount of time between these two experiments. I believe this was an adequate way to address any concerns about the first experiment influencing the results of the second. Another notable difference between the two experiments was that in providing the social comparisons in *Experiment 3,* I only indicated if the school was in the top- or bottom-half of local schools. That is, I did not include a rank (e.g., 3rd out of 10). This omission was deliberate and an attempt to help respondents not conflate the two experiments despite their similarities. It also allowed me to address a potential concern about causal inference. Specifically, if a respondent sees a prompt that reads "top half", we wouldn't know if she interprets this as first out of three or 49th out of 100.

Experiment 2[5]

In Experiment 2 I asked respondents to rate the performance of an unnamed high school (High School A) using performance data from a standardized English exam. The goal was to observe the assessed performance when both historical and social comparison information were presented together. I felt this would be a suitable way to design the experiment for two reasons. First, in a realistic organizational decision-making environment (i.e., a non-experimental setting), managers might have a sense of their organization's performance as well as the performance of peer and competitor organizations. Second, by including both comparison types in the same experimental frame we might be able to get some sense of the strength of the positive and negative versions of each comparison. Of course, I was also able to compare performance assessments against the control group as well.

---

[5] Here, I only show the "awareness" measure in the results section. The empirical findings for these experiments regarding the interpretation component of the framework can be found in Essay 3 of this dissertation. Of note, the findings on the interpretation part of this experiment indicated the presence of cognitive bias in the use of performance information.

In this experiment, each individual (regardless of treatment group assignment) saw a raw performance metric which stated that 77% of students at High School A passed the English exam. Respondents were randomly assigned to one of five groups. The control group saw only the raw performance metric. The other groups saw four combinations of historical and social comparisons. The historical comparison prompts said that the performance was indicative of a 2% increase or decrease in the rate of students who passed the standardized English exam. The social comparison indicated that based upon the pass rate that the school ranked third or seventh out of ten comparable local schools. For the social comparison prompt, individuals were told if this was in the top- or bottom-half of local schools, respectively. Individuals were then asked to rate the performance of the school using a 0-100 sliding scale.

As an example, someone in the group that saw prompts indicating increases for both the historical and social comparisons saw the following prompt:

> "English Exam: 77% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10). Assuming this is the only information available to you, use the sliding scale (0-100) to assess the overall performance of HIGH SCHOOL A over the last year:"

Respondents would rate the performance of the school with the sliding scale. At the end of the survey respondents were asked to recall the value of the performance metric they saw in this experiment. The value they provided is the dependent variable of interest in this study for Experiment 2.

Experiment 3

32

In Experiment 3 I asked respondents to rate the performance of an unnamed high school (High School B) using performance data from a standardized Math exam. In Experiment 3, respondents only saw one comparison at a time so that we could get a sense of the strength of the comparisons by themselves in the assessment of performance data. I used similar comparisons from Experiment 2. Again, respondents were randomly assigned to one of five groups.

Individuals saw a raw performance metric that stated that 79% of students at this high school passed the Math exam. As before, the control group saw only the raw performance metric. The other groups saw one of four possible historical and social comparisons. That is, groups 2-5 only saw one of the following: 2% increase from last year, 2% decrease from last year, top-half of comparable local schools, or bottom-half of comparable local schools. Again, respondents were asked to rate the performance of the school on a 101-point sliding scale. Finally, at the end of the survey, respondents were asked to report the raw value of the performance metric they observed (79).

Results

Table 6 provides the by group mean response to the observed performance metric respondents saw for both experiments.

Table 6 - Group Means by Treatment Group			
English Exam		Math Exam	
Treatment Group	Mean	Treatment Group	Mean
Control	75.83	Control	78.84
Increase (H) Top Half (S)	76.05	Increase (H)	77.47
Increase (H) Bottom Half (S)	73.69	Decrease	77.71
Decrease (H) Top Half (S)	74.39	Top Half (S)	78.24
Decrease (H) Bottom Half (S)	76.00	Bottom Half (S)	76.78

The next two tables (7 & 8) show the percentage of respondents in each group who successfully recalled the observed performance metric. In Experiment 1, one-third of respondents correctly recalled the raw performance metric. One can observe that these values are much larger. This is possibly a result of the fact that the performance metric was randomized in Experiment 1 but fixed for these two experiments.

Table 7 - Respondents Who Accurately Recalled the Observed Performance Metric			
English Exam			
Treatment Group	Frequency	N	Percent
Control	52	59	88.14%
Increase (H) Top Half (S)	46	57	80.70%
Increase (H) Bottom Half (S)	51	65	78.46%
Decrease (H) Top Half (S)	51	62	82.26%
Decrease (H) Bottom Half (S)	47	57	82.46%

34

Table 8 - Respondents Who Accurately Recalled the Observed Performance Metric			
Math Exam			
Treatment Group	Frequency	N	Percent
Control	52	63	82.54%
Increase (H)	51	62	82.26%
Decrease	40	58	68.97%
Top Half (S)	41	58	70.69%
Bottom Half (S)	40	59	67.80%

For responses to the English Exam experiment, I conducted a one-way ANOVA to determine if the stated value of the observed performance metric (77) was different across five (a control and four treatment) groups. There was not a statistically significant difference in the reported value of the observed performance metric across these groups as determined by one-way ANOVA ($F(4,295) = 0.61$, $p = 0.65$). This did not pass Bartlett's test for equal variances $\chi^2(4) = 54.74$, $p = 0.00$, suggesting the need to run a Kruskal-Wallis equality-of-populations rank test. The results from that test also did not suggest a statistically significant difference in the stated value of the observed performance metric, $\chi^2(4) = 0.71$, $p = 0.95$.

For responses to the Math Exam experiment, I conducted a one-way ANOVA to determine if the stated value of the observed performance metric (79) was different across five (a control and four treatment) groups. There was not a statistically significant difference in the reported value of the observed performance metric across these groups as determined by one-way ANOVA ($F(4,295) = 0.89$, $p = 0.47$). This did not pass Bartlett's test for equal variances $\chi^2(4) = 25.64$, $p = 0.00$, suggesting the need to run a Kruskal-Wallis equality-of-populations rank test. The results from that test also did not suggest a statistically significant difference in the stated value of the observed performance metric, $\chi^2(4) = 2.62$, $p = 0.62$.

35

Note, the above only show that group randomization did not affect the post-hoc check. A tougher test looks to see if the stated performance outcome may have influenced the value respondents indicated they saw. Because respondents all saw the same metric we cannot include the observed value in these regressions (as in the previous experiment). To dig deeper, we want to see if how individuals interpret the outcome might influence the value they provide in the "check" question (this is a harder test). Table 9 shows whether the outcome score or outcome score and treatment group influenced the awareness check for either experiment.

Variable	English Exam		Math Exam	
	Model 1	Model 2	Model 3	Model 4
Outcome Score	0.09**	0.09**	0.04	0.03
	(2.46)	(2.44)	(1.45)	(1.2)
Treatment Group		0.07		-0.27
		(0.15)		(-1.01)
Constant	69.23***	69.02***	75.11***	76.05***
	(27.84)	(24.27)	(39.42)	(35.86)
N	300	300	300	300
Prob > F	0.0146	0.0504	0.1491	0.2128
R-squared	0.0198	0.0199	0.007	0.0104

Table 9 - OLS Regressions

Dependent variables are the responses to solicitations for the observed performance metric (77 for English Exam, 79 for Math Exam). ** $p < 0.05$, *** $p < 0.01$

Of note, the outcome score and the treatment group did not meaningfully influence responses to the Math Exam. The outcome score was significant for responses to this part of the English Exam but not the treatment group assignment. This latter point reflects the findings from the ANOVA test. But, it is probably the case that including the observed value in the regressions for the English exam (if this were possible) would make the significant relationship of the outcome score on the stated observed value go away. Nonetheless, it's interesting that the

outcome score is statistically significant for one experiment but not the other even though the two experiments are strikingly similar.

Discussion

The three experiments in this study are meant to open our understanding of where it is in the cognitive process of performance information use that cognitive bias may influence the way that public managers use performance information. To my knowledge, this is the first manuscript in public management to attempt this.

Previous scholarship on performance information use by public managers can be characterized to fall into two camps. A majority of this work adopts an organizational theory perspective and is highlighted by Moynihan's Interactive Dialogue Model, one of the leading models that scholars use to think about how public managers use performance information. Another growing body of scholarship highlights the role of psychology in explaining how public managers use performance information. Unfortunately, to date, little scholarship has attempted to speak across these two bodies of work.

While acknowledging the value of the IDM, I use recent empirical evidence to suggest some shortcomings in the assumptions of the model. Specifically, I suggest the model should be updated to consider how public managers *process* performance information. Following a framework of political information use put forward by Gaines and colleagues (2007), I offer the Performance Information Processing Framework (PIPF). In this framework, the acts of being "aware" of performance information—that is, accurately knowing what the information is or says—and interpreting the performance information, represent distinct steps in which cognitive biases may influence the way that individuals (i.e., public managers) process and eventually use

37

performance information. Following this framework, cognitive processing might influence the way that public managers use performance information. Accordingly, my expectation was that the act of interpreting performance information might be prone to the influence of cognitive bias.

In three experiments that investigated the level of awareness that respondents had of the raw performance metric they previously viewed in an experimental treatment, I found considerable evidence that respondents were equally "aware" of the information they viewed. While there was variation in the ability of respondents to accurately recall the exact performance metric they saw across the three experiments, I found little evidence of any statistically significant relationship between a respondent's treatment condition and their response when requested to recall the performance information in question. This suggests that framing manipulations that were intended to elicit various cognitive biases have little influence on a respondent's ability to know what the information is. These results provide experimental evidence in support of part of this framework. This can help management scholars explain how and why public managers use performance information.

Further, any evidence for variation in the interpretations of the performance information observed in these experiments across experimental frames should point to the process of interpretation as being the step in the cognitive processing of performance information in which individuals are most likely to exhibit cognitive biases. (**Note, this is exactly what I demonstrate in the remainder of this dissertation. Thus, in line with the findings from Gaines and colleagues, experimental evidence suggests that Model 3, *meaning avoidance*, represents the most accurate model of the cognitive processes a manager undergoes when using performance information**.)

38

The PIPF and these findings suggest the importance of incorporating behavioral explanations into our frameworks and models of how we conceptualize how individuals interact with performance information. Also, this work acknowledges the potential for other and future studies to explore what it means to "use" performance information. To expand on these points, a behavioral perspective potentially suggests that performance information use means something different than has been previously considered in the literature. One insight this perspective might provide to public management scholars, for example, is that researchers would need to be more deliberate in how they theorize about the act of performance information use. Is there a direct link between interpretation and the use stage, as the PIPF suggests? Does this depend on what type of use—i.e., persuasion or making financial decisions for one's organization—might be of interest?

Limitations

In addition to the contributions of this study, there are some limitations which merit further discussion. To begin, as this is experimental work, the usual caveats about questions of how well this explains what happens in practice apply. If we accept the PIPF and the approach undertaken herein, it is also worth mentioning how future scholarship might be able to extend some of the theoretical limitations of this body of scholarship.

First, how well does the PIPF capture moving from interpretation to use? It could be that there are other cognitive stages that might influence how public managers use performance information. This should be explored in future research. Second, how might incorporating organizational theory influence the PIPF? In addition to the previous point, it could very well be that institutional factors in the organizational or political environment (including and in addition

39

to one's organizational role) could influence how individuals process information under certain conditions. Or, it might be that organizational factors influence what steps in the PIPF are relevant to the circumstances of a public manager's decision making process beyond an experimental setting. This logic is already deeply rooted in the IDM.

Third, and relatedly, how well can the framework be incorporated into the Interactive Dialogue Model? The primary critique of the IDM this study raised was that despite the contributions and value of the IDM, there remain limitations to the model. I believe the experimental findings presented here will be of interest to scholars working in this area because of how the findings point to the value of looking for behavioral mechanisms of our empirical findings. Yet, how well does the PIPF fit into the IDM? Or, should scholars consider it as a standalone framework with four distinct models that potentially describe performance information use under different conditions? I personally feel the IDM is robust enough to accommodate the few changes in assumptions necessary to integrate the PIPF into the larger IDM framework. To this point, I could have offered up the PIPF as a distinct perspective. While this may have led to some amount of scholarship interested in this perspective, I felt it was antithetical to the larger intellectual project of trying to understand how public managers *use* performance information. In this sense, I think trying to incorporate the PIPF into the IDM offers more to scholars over the long run.

Conclusion

In recent years, public management scholars have given significant attention to the question of how individuals, including public managers, use performance information. Despite some theoretical markers for how to think about this area of research, more effort has been given

40

to describing the subject through empirical work. For this reason, there is a need for public management scholars to seek to develop our understanding of how individuals use performance information in a way that combines theoretical and empirical work from the past two decades. This essay seeks to do this by combining the Interactive Dialogue Model as a framework for thinking about how public managers use performance information and the recent empirical work that demonstrates the behavioral foundations of performance information use. I present a framework—the performance information processing framework (PIPF)—that seeks to describe the cognitive process of performance information use over four distinct cognitive processing models. I believe this framework can be fully incorporated into the Interactive Dialogue Model with only slight modifications to the assumptions of the IDM.

A growing body of empirical evidence suggests cognitive biases can play a significant role in how individuals use performance information. Yet, heretofore, we have had a limited understanding of a) why that is, b) when it occurs, or c) how we might be able to design performance management systems in a way that mitigates these biases. The PIPF and the empirical results found in this essay help to uncover some of those mysteries. Specifically, I show that individuals, at least in an experimental setting, show similarities in understanding what the performance information they see says. Thus, any evidence for differences in assessments of performance information will likely come from the process of interpreting the raw performance information. That is, cognitive bias is most likely to influence individuals not in their ability to know what performance information is, but in their ability to interpret what it says.

Finally, this research should speak to students of behavioral public administration in several ways. First, it should highlight the need and value in looking at behavioral mechanisms as framework for thinking about how to incorporate psychological insights into public

41

management scholarship. Second, it suggests that theorizing about behavioral constructs and processes may offer a bevy of opportunities to better understand our phenomena of interest. Third, when it comes to public management, more effort is needed in understanding how the interplay of behavioral and organizational theories matters for managing in the public sector.

References

Ammons, D. N. and W. C. Rivenbark (2008). "Factors influencing the use of performance data to improve municipal services: Evidence from the North Carolina benchmarking project." <u>Public Administration Review</u> 68(2): 304-318.

Andersen, S. C. and M. Hjortskov (2015). "Cognitive Biases in Performance Evaluations." <u>Journal of Public Administration Research and Theory</u> 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." <u>Journal of Public Administration Research and Theory</u>: 1-13.

Anderson, D. M. and B. C. Edwards (2015). "Unfulfilled Promise: Laboratory experiments in public management research." <u>Public Management Review</u> 17(10): 1518-1542.

Baekgaard, M., J. Christensen, C. M. Dahlmann, A. Mathiasen and N. B. G. Petersen (2017). "The role of evidence in politics: Motivated reasoning and persuasion among politicians." <u>British Journal of Political Science</u>: 1-24.

Baekgaard, M. and S. Serritzlew (2016). "Interpreting performance information: Motivated reasoning or unbiased comprehension." <u>Public Administration Review</u> 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." <u>Journal of Public Administration Research and Theory</u>: muw028.

Bartels, L. M. (2002). "Beyond the running tally: Partisan bias in political perceptions." <u>Political Behavior</u> 24(2): 117-150.

Behn, R. D. (1995). "The big questions of public management." <u>Public Administration Review</u>: 313-324.

Behn, R. D. (2003). "Why measure performance? Different purposes require different measures." <u>Public Administration Review</u> 63(5): 586-606.

Bouwman, R. and S. Grimmelikhuijsen (2016). "Experimental public administration from 1992 to 2014: A systematic literature review and ways forward." International Journal of Public Sector Management 29(2): 110-131.

Charbonneau, É. and G. G. Van Ryzin (2015). "Benchmarks and citizen judgments of local government performance: Findings from a survey experiment." Public Management Review 17(2): 288-304.

Christensen, J., C. M. Dahlmann, A. H. Mathiasen, D. P. Moynihan and N. B. G. Petersen (2018). "How do elected officials evaluate performance? Goal preferences, governance preferences, and the process of goal reprioritization." Journal of Public Administration Research and Theory 28(2): 197-211.

Cox, C. (1999). Nietzsche: Naturalism and Interpretation. Berkeley, CA, University of California Press.

Epley, N. and T. Gilovich (2016). "The mechanics of motivated reasoning." Journal of Economic Perspectives 30(3): 133-140.

Gaines, B. J., J. H. Kuklinski, P. J. Quirk, B. Peyton and J. Verkuilen (2007). "Same facts, different interpretations: Partisan motivation and opinion on Iraq." Journal of Politics 69(4): 957-974.

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.

Gerber, A. and D. Green (1999). "Misperceptions about perceptual bias." Annual Review of Political Science 2(1): 189-210.

Gerber, A. and D. P. Green (1998). "Rational learning and partisan attitudes." American Journal of Political Science: 794-818.

Gigerenzer, G. and W. Gaissmaier (2011). "Heuristic decision making." Annual review of psychology 62: 451-482.

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

44

Harvey, D. (2001). <u>Spaces of Capital: Towards a Critical Geography</u>. New York, Routledge.

Hochschild, J. L. and K. L. Einstein (2015). "Do facts matter? Information and misinformation in American politics." <u>Political Science Quarterly</u> 130(4): 585-624.

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

James, O., S. R. Jilke and G. G. V. Ryzin, Eds. (2017). <u>Experiments in Public Management Research: Challenges and Contributions</u>. New York, Cambridge University Press.

James, O. and G. G. Van Ryzin (2016). "Motivated reasoning about public performance: An experimental study of how citizens judge the affordable care act." <u>Journal of Public Administration Research and Theory</u> 27(1): 197-209.

Jilke, S., S. Van de Walle and S. Kim (2016). "Generating usable knowledge through an experimental approach to public administration." <u>Public Administration Review</u> 76(1): 69-72.

Kroll, A. (2015). "Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research." <u>Public Performance & Management Review</u> 38(3): 459-486.

Kroll, A. (2015). "Explaining the Use of Performance Information by Public Managers: A planned-behavior approach." <u>The American Review of Public Administration</u>: 201-215.

Kunda, Z. (1990). "The Case for Motivated Reasoning." <u>Psychological Bulletin</u> 108(3): 480-498.

Litman, L., J. Robinson and T. Abberbock (2016). "TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences." <u>Behavior research methods</u>: 1-10.

Meier, K. J., N. Favero and L. Zhu (2015). "Performance gaps and managerial decisions: A Bayesian decision theory of managerial action." <u>Journal of Public Administration Research and Theory</u> 25(4): 1221-1246.

Moynihan, D. P. (2008). <u>The Dynamics of Performance Management: Constructing Information and Reform</u>. Washington, DC, Georgetown University Press.

Moynihan, D. P. (2010). The Promises and Paradoxes of Performance-Based Bureaucracy. <u>The Oxford Handbook of American Bureaucracy</u>. 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., P. A. Nielsen and A. Kroll (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. Emperiments in Public Management Research: Challenges and Contributions. O. James, S. R. Jilke and G. G. V. Ryzin. New York, Cambridge University Press: 244-269.

Moynihan, D. P., P. A. Nielsen and A. Kroll (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. Experiments in Public Management Research: Challenges and Contributions. O. James, S. R. Jilke and G. G. V. Ryzin. New York, Cambridge University Press: 244-269.

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 20(4): 849-866.

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.

Nyhan, B. and J. Reifler (2010). "When corrections fail: The persistence of political misperceptions." Political Behavior 32(2): 303-330.

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. (2015). "Citizen (dis) satisfaction: An experimental equivalence framing study." Public Administration Review 75(3): 469-478.

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

Olsen, A. L. (2017). "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." Journal of Public Administration Research and Theory.

Oppenheimer, D. M. and E. Kelso (2015). "Information processing as a paradigm for decision making." Annual review of psychology 66: 277-294.

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.

Schaffner, B. F. and C. Roche (2016). "Misinformation and motivated reasoning: Responses to economic news in a politicized environment." Public opinion quarterly 81(1): 86-110.

Taber, C. S. and M. Lodge (2006). "Motivated skepticism in the evaluation of political beliefs." American Journal of Political Science 50(3): 755-769.

Van de Walle, S. and W. Van Dooren (2011). Introduction: Using Public Sector Performance Information. Performance Information in the Public Sector: How it is Used. W. Van Dooren and S. Van de Walle. New York, Palgrave Macmillan: 1-12.

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.

Yeo, S. K., M. A. Cacciatore and D. A. Scheufele (2015). News selectivity and beyond: Motivated reasoning in a changing media environment. Publizistik und gesellschaftliche Verantwortung, Springer: 83-104.

Yockey, M. D. and S. M. Kruml (2009). "Everything is relative, but relative to what? Defining and identifying reference points." Journal of Business and Management 15(1): 95.

47

Framing and Benchmarks in the Use of Performance Information:
Two experiments on the role of cognitive bias in the use of performance information

Sean Webeck
Indiana University

Abstract: Public sector performance management systems, and the performance information they create, have received a lot of scholarly attention. Early work suggested the role of institutions in shaping how public managers use performance information. More recent scholarship suggests that individuals exhibit cognitive biases in the use of performance information. This essay contributes to this literature and suggests behavioral and institutional explanations for how public managers interpret performance information. This extends current work in several important ways. Through survey experiments, I provide evidence on the way individuals familiar with performance information process this information. Two experiments— one fielded in a sample of professional managers and another fielded in individuals with significant public sector work experience—suggest cognitive bias plays an important role in the way in which public officials process performance information. I also incorporate two administrative institutions—performance benchmarks and decision justification requirements— into the second experiment. Results suggest performance benchmarks play a significant and substantively large role in how individuals process performance information. Taken together, these findings suggest understanding how public managers process and interpret performance information in the context of administrative institutions provides several avenues for future research.

48

Introduction

As Gormley and Balla note, "In the 1990s the concept of **performance** came to rival accountability as a standard for evaluating executive branch agencies" (Gormley and Balla 2008 [14, emphasis in original]). Over the years, a growth in the belief that democratic societies could improve public sector performance by focusing on this concept led many public organizations to adopt performance management systems. At the heart of these systems lies a belief that quantitative measures of performance represent an important instrument to better manage public sector organizations. As one scholar noted, "the dissemination of quantitative measures of performance has been one of the most widespread trends in government in past decades" (Moynihan 2015, 33). Unfortunately, our understanding of how public managers use the information created by these systems has not kept pace with their rapid diffusion.

Scholarship on the question of how public managers use performance information falls into two camps. The first writings addressed the roll of institutions in shaping information use. A recent review assesses the progress made in this area but highlights areas of future research (Kroll). A more recent body of scholarship addresses one of the questions Kroll highlighted; this work seeks to understand how psychology influences the use of performance information in public managers. These studies align with studies of performance information use by other users of interest—i.e., citizens and political officials—and fits within a recent call for scholars of public administration to more deliberately incorporate the role of micro-level processes into their work (Grimmelikhuijsen et al. 2016). Somewhat ironically, while a majority of the work in this area can be categorized to fall within one of the two areas, one of the leading theories for how public managers use performance information attempts to explain this behavior by incorporating both perspectives—the behavioral and the institutional (Moynihan 2008).

49

Here, I seek to contribute to our understanding of how scholarship on performance information use by public managers should incorporate both individual- and institutional-level explanations for this behavior. I argue scholars should study this topic within the larger framework of decision making. A decision making framework encourages us to consider decision making as a process rather than as a discrete outcome. While it has been acknowledged within the study of performance information use (Moynihan 2008), previous work on the topic largely ignores this approach. Nonetheless, scholarship on cognitive biases speaks to way in which psychological foundations can influence the way people consider information and, more broadly, make decisions. Also, situated within a broader literature on institutions, work on accountability instruments—e.g., performance benchmarks and requirements to justify decisions—provides evidence that institutions influence the way human beings interact with their environment.

Empirically, I present evidence in support of these arguments from two different experiments. I fielded Experiment 1 in a group of professional managers (150 private and 150 public sector managers). For Experiment 2, I fielded a similar experiment through a survey of public sector employees and individuals with significant experience working in the public sector drawn from MTURK. The use of online survey experiments is in line with a recent discussion of the value in using an online web service to collect data for public management research (Stritch et al. 2017). The findings from these experiments provide further evidence for the role of cognitive bias in the interpretation of performance information. Additionally, it suggests performance benchmarks serve as important reference points in the use of performance information. These results give credence to the claim that future research should give greater

attention to understanding the nexus of institutions and individual-level micro-processes—both in a practical and a theoretical sense—in bureaucratic decision making.

The remainder of the paper will proceed as follows. In the next section, I build an argument for the value in trying to integrate macro- and micro-level processes in understanding performance information processing. I then focus attention on framing effects and reference points, suggesting why these factors might influence how public managers use performance information. Next, I discuss two institutions that could play a role in the interpretation of performance information. The empirical section describes the research design of both experiments before addressing the results of each experiment. I end with a discussion followed by my concluding remarks.

Literature
Performance Management Systems and the Study of Performance Information
Throughout the performance management movement, leading scholars have tried to direct attention to the question of how public managers use performance information (Behn 1995, Moynihan and Pandey 2010). Yet, recently, a prominent scholar commented that "we know little about the basic tendency of individuals to incorporate and use performance information" (Moynihan 2015). Moynihan defines performance management as "a system that generates performance information through strategic planning and performance measurement routines and that connects this information to decision venues, where, ideally, the information influences a range of possible decisions" (2008, 5). Thus, performance management systems are decision making systems. Accordingly, to understand how public managers use performance information, which is arguably among one of the most important unanswered questions regarding the study of public organizations, we must understand how they *make decisions* about performance.

51

I categorize our lack of understanding on this issue into three classes of assumptions. The first involves a set of assumptions we make about performance management systems and the information they create. Two that stand out as relevant to the current study are: 1) "Government can and should make more rational decisions", and 2) "Performance information will improve decisions and can be used to foster accountability" (Moynihan 2008, 27). The link between performance information and accountability through the quality of decisions made based upon this information rests on a further assumption about the information itself. Performance data were originally thought to allow for an individual to make an objective assessment of how an organization is doing because it is intended to be "systematic" information (Radin 2006, Nielsen 2013). It's often numerical format provides a "reassuring status of clarity and objectivity" (Moynihan 2008, 95). The promise of this story has led us to a point in which performance management systems are ubiquitous. Nonetheless, while performance management doctrine assumes bureaucrats will incorporate the performance information these systems create into their decision making, at this point there is sufficient empirical evidence (reviewed below) to refute the basic assumptions listed here.

The second class of assumptions involves how we conceptualize how individuals respond to information and make decisions. That the study of human decision making has received a lot of scholarly attention is not surprising because how human beings use information to make decisions is one of the most important questions in the social sciences (Giles 2011). One consequences of this is the ability to broadly define studies of human decision making as falling within one of two camps. The first second focuses on institutions, demonstrating that under many conditions, humans respond in predictable ways to these institutions. The second centers on the role of psychology in understanding how individuals process information.

52

A related, but third, assumption centers on what it means to "use" performance information. Previous research seems to characterize use as a discrete act that follows from observing a performance metric (Behn 2003, Moynihan 2010, Van de Walle and Van Dooren 2011, Van Dooren and Van de Walle 2011). This is a critical assumption. One potential shortcoming is the way in which what happens at the micro-level is unobservable. An information processing approach gives scholars more theoretical leverage. It allows us to observe how individuals respond to performance information over multiple cognitive steps. For example, it allows us to observe whether or not individuals are actually "aware" of the information in question (and if so, how well) and, if so, how they "interpret" the information (Gaines et al. 2007).

If we are to accept the claim that how bureaucrats use performance information is an important question for public management researchers (Moynihan and Pandey 2010), we must acknowledge some of the limitations of previous research to theorize and teach future practitioners about the subject (Kroll 2015, Moynihan et al. 2017). Specifically, progress thus far considers macro- and micro-level processes independent of one another. That is, studies of performance information use by public managers either incorporate an institutional or psychological perspective but not both. This is in line with a critique from Priem and colleagues about management studies, broadly construed. They suggested that "those individual judgment studies that have been performed by management researchers have almost always stayed within one level of analysis" (Priem et al. 2011, 554-555). One assumption advanced here is that to make progress on the basic question for performance information use by public managers, we must figure out how to integrate both the macro- and micro-level processes that work to shape the phenomenon.

53

Institutions and Performance Information Use

Early studies on public managers focused on the question of how aspects of the institutional environment drive performance information use. A recent systematic literature review of the use of performance information outlined some of the progress made in this area since 2000 (Kroll 2015). Over the last 15 years in "a highly relevant and fast-growing research area" (ibid., 460), research consistently shows six factors commonly drive the use of performance information among bureaucrats: 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), 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. 2012), and goal clarity (for examples see Moynihan and Landuyt 2009, Moynihan et al. 2012, Moynihan et al. 2012). Thus, previous scholarship suggests that, as expected by institutional theory, the involvement of external stakeholders influences if and how bureaucrats use performance information. Relevant to this study, a majority of previous studies focus on the ways in which the organizational context in which performance measurement systems are embedded influence on how the information from those systems gets used.

Behavior and Performance Information Use

More recently, some scholars have pointed to behavioral factors that might produce systematic variation in the use of performance information among bureaucrats (Kroll 2015,

Moynihan et al. 2017). These, and other studies, suggest psychology may play a role in helping us understand how public managers use performance information (Moynihan 2008, Salge 2011, Nielsen 2013, Kroll 2015, Moynihan 2015, Andersen and Moynihan 2016). In addition to these, some very recent pieces also demonstrate the utility of incorporating an individual-level behavioral approach to examine the use of performance information. Other studies suggest this perspective can contribute to our understanding of how individuals, broadly considered, respond to performance information. These include studies on citizens (Olsen 2013, Andersen and Hjortskov 2015, Olsen 2015, Baekgaard and Serritzlew 2016, Barrows et al. 2016, Hvidman and Andersen 2016, Olsen 2017) and politicians (Olsen 2014, Nielsen and Baekgaard 2015, George et al. 2016, Nielsen and Moynihan 2016). This approach should be understood within a broader context of seeking to understand the psychological foundations of public administration (Stephen Grimm).

Institutions, Behavior, and Performance Information Use
While both perspectives have contributed to our understanding of how individuals—including public managers—use performance information, scholars of public management have yet to undertake significant efforts to try and synthesize these two perspectives. One attempt that stands out is the Interactive Dialogue Model (IDM) (Moynihan 2008). This model describes how and why public managers use performance information. The important takeaway from the model is that performance information can be subjective. According to the IDM, performance information is ambiguous because of political considerations *a priori* to any descriptive story of performance information "use". This idea of subjectivity is contrary to the performance management doctrine, representing a major break in the theoretical development of performance information use. In Moynihan's model, "political considerations" are taken to represent

55

motivated reasoning as a cognitive bias that adds subjectivity. This bias arises out of an individual's role within an organization or larger political context.

I suggest two limitations to the IDM. First, more recent empirical evidence suggests other cognitive biases might influence the use of performance information. Second, there is an overreliance on role as an institutional variable. If we understand that the fundamental problem in public administration is that bureaucrats make public decisions that have public consequences, we must recognize that citizens and political officials use a multitude of tools to try use to constrain bureaucratic behavior (Bertelli and Lynn 2006, Meier and Bohte 2007). While the IDM appears to offer the flexibility necessary to accommodate revisions in its assumptions of the role of behavior and institutions in how public managers use performance information, the model presently stands in need of a revision that offers a broader array of variables for public management scholars to consider. Here, I offer aspects of behavior—reference points, framing effects, and negativity bias—and institutions—performance benchmarks and justification requirements—as elements that might influence how public managers interpret performance information.

Theory
General Expectations
In the Interactive Dialogue Model, subjectivity arises from a cognitive bias—motivated reasoning—that arises from a public manager's role within a public sector organization. Here, I argue that public managers may exhibit cognitive biases—e.g., framing effects or responses to reference points—simply because of the way information is presented to them. I also offer that certain accountability tools—performance benchmarks and decision justification requirements— should moderate these biases under appropriate conditions. In short, public managers may

interpret ostensibly objective performance information in a subjective way that simply reflects the use of different points of reference.

The labels "System 1" and "System 2" are used to describe two very different cognitive processes (Kahneman 2011). System 1 processing reflects an unconscious action where human beings use heuristics to think fast, make many associations, and generally process as much information as possible. System 2 processing, on the other hand, is more deliberate, more "rational". In this mode, individuals undertake more time and effort to consciously undertake a more reliable decision making process. While the IDM focuses on motivated reasoning, other research on the use of performance information suggests that cognitive bias can be observed when individuals exhibit System 1 processing.

In a broad sense, I expect that when presented with situations that reflect the biases discussed herein, public managers will tend to exhibit System 1 thinking. That is, public managers will respond to framing effects in a way that reflects System 1 thinking. I believe these effects will increase in the context of a performance benchmark but decrease when public managers are asked to provide a justification of their decisions. Evidence for these biases will come when public managers respond to qualitatively equivalent information in substantively different ways. That is, a cognitive bias will be present when public managers *interpret* the same performance information in different ways when the only difference is the way that information is presented to them.

As a caveat, these expectations may depend on the scale of a comparison being made or some other artifact of the interpretative process. That is, the hypothesis in this type of behavioral research is not that everyone will exhibit the cognitive bias in question but that a statistically distinct percentage of individuals will exhibit the bias.

57

Reference Points and Cognitive Bias

The preceding discussion relies heavily on the idea that reference points will serve as a significant component of the way that public managers will interpret performance information. In the following discussion, I expect reference points will help to facilitate System 1 thinking and cognitive bias because individuals will consider the performance metric in the context of some reference point rather than as an objective measure.

Reference Points

Reference points are "stimuli of known attributes that act as standards against which other categorically similar stimuli of unknown attributes are compared in order to gain information" (Yockey and Kruml 2009, 97). Reference points represent a significant part of our cognitive processing because our judgment is fundamentally comparative in nature (Mussweiler 2003). That is due to the fact that our perception is "*reference-dependent*" (Kahneman 2002, 459, emphasis in original). And, the way we fixate on reference points tends to add a level of subjectivity to the way we interpret decisions and events around us.

Some argue that public managers will use comparisons in how they think about performance (Meier et al. 2015, Olsen 2015). Others suggest there is some evidence to suggest of this (Askim et al. 2007, Ammons and Rivenbark 2008, Hammerschmid et al. 2013, Nielsen 2014). Nonetheless, questions pertaining to the importance of reference points in how bureaucrats use performance information remain largely unexplored.

Attribute Framing

How individuals respond to framing effects represents another cognitive bias. The study of framing effects suggests "decision makers respond differently to different but objectively equivalent descriptions of the same problem" (Levin et al. 1998). Levin and colleagues suggest there are three types of frames: risky choice framing, attribute framing, and goal framing. This

58

study uses attribute framing to assess how bureaucrats respond to performance information with negative or positive frames. Attribute framing involves changing the frame used to understand a key attribute, such as rates of satisfaction/dissatisfaction, success/failure, or employment/unemployment (Levin et al. 2002, Olsen 2015, Olsen 2015).

Since prior preferences can be used "to evaluate the impact of a frame on unadulterated preferences" (Druckman 2001), one way to explore the strength of response to performance information is to compare how individuals respond to performance information compared to how individuals in a baseline categories of reference responded to similar information. The literature on performance information use suggests individuals do respond to framing effects when considering performance information (Charbonneau and Van Ryzin 2015, Olsen 2015). While there is some evidence to these questions, the magnitude of a frame that is necessary to move someone out of the margin of indifference and the magnitude of response remain underexplored.

Negativity Bias

The discussion of the strength of a response to a frame naturally leads to the idea of a negativity bias. It is well-known that individuals respond more strongly to negative information than to objectively equivalent information framed in the positive (Ito et al. 1998, Baumeister et al. 2001, Rozin and Royzman 2001). The general phenomenon is known as negativity bias, when "negative events are more salient, potent, dominant in combinations, and efficacious than positive events" (Rozin and Royzman 2001, 297). They continue:

> "The principle of negative potency asserts that, given inverse negative and positive events of equal objective magnitude, the negative event is subjectively more potent and of higher salience than its positive counterpart. More generally, the claim is that negative events are more potent with respect to their objective magnitude than are positive events" (Rozin and Royzman 2001, 298).

59

Through the importance of (dis)satisfaction with public sector organizations (James and John 2007, Charbonneau and Van Ryzin 2015, Olsen 2015, James and Van Ryzin 2017), negative potency is one form of negativity bias that is especially relevant to the study of performance information use. But, it is unclear if public managers should exhibit this bias as do citizens.

Institutional Activators of Cognitive Bias

My expectation is that the influence of these well-established cognitive biases on the interpretation of performance information can itself be influences by institutional (i.e., micro) characteristics. We know that organizations do not present information in a vacuum, but instead require that managers use performance information to make certain comparisons with that information. Very often those comparisons are dictated by previously determined levels of acceptable performance, which are determined by levels of performance in similar organizations. These types of comparison are commonly known as benchmarking. In addition to benchmarking, some performance management systems require managers to justify their interpretation of and reactions to performance information. These accountability systems are meant to ensure productive feedback processes and allow agency leaders and/or political principals to monitor and better understand the use of performance information. This section will briefly discuss the literature on the potential impacts of these institutional features on type 1 vs. type 2 thinking in the use of performance information.

Benchmarking

Simply stated, benchmarking is, "The process of comparing performance across organizations" (Bouckaert and Van Dooren 2009 156). It comes in two forms (Löffler 2001). Absolute benchmarking occurs when predefined standards of performance lead to a "pass-fail" approach to understanding organizational performance. In this way, any organization (or

60

individual) can "pass" a benchmark. On the other hand, relative benchmarks require competition between comparable organizations. A common way to institute a relative benchmark is to allow for a fixed number of winners who have the best performance among the group in question. This type of reference point would qualify as "goal framing" (rather than attribute framing) in Levin and colleagues' categories of frames (Levin et al. 1998).

Justifications

It is well known that while political officials need administrators to undertake actions on their behalf, they are motivated to constrain bureaucratic behavior towards producing a set of decisions and results that are more favorable to their preferences (West 1995, Seidenfeld 1996, Seidenfeld 2001, West 2004). One tool that is commonly used to constrain bureaucratic behavior is the need to explain—to justify—*ex-post*, the motivations and reasoning behind a particular bureaucratic action. This has been said to affect the underlying psychology of bureaucratic behavior (i.e., decision making) because it "encourages agencies to take greater care when formulating rules, which in turn decreases the likelihood that the rulemaking process will reflect psychological decisionmaking [*sic*] biases" (Seidenfeld 2001, 1060).

Hypotheses

Based on the theory outlined above, this section lays out explicit hypotheses regarding individual and institutional influences on the interpretation of performance information.

Framing Effects and Negativity Bias

I expect that public managers will respond to framing effects. That is, when qualitatively similar information is framed to one group in the positive and another in the negative, the group who sees the information framed in a negative way will provide a lower performance assessment (i.e., their intereptation of the raw data will be different) than the group that saw the positive frame.

61

As a corollary, I expect that responses will be stronger when information is observed in a negative frame than to equivalent information framed in the positive. That is, I expect the absolute difference in change between a response and a baseline reference category will be greater for those who see negative information than for those who see qualitatively similar information in a positive frame. Both of these expectations are consistent with a great volume of experimental evidence other decision contexts.

Performance Benchmarks

I expect that performance benchmarks will set the reference points that public managers use to determine if current performance is above or below acceptable levels. That is, even though benchmarks often represent a type of institution designed to increase accountability in line with the assumptions of the performance management doctrine, the fact that benchmarks lead individuals to undergo the same type of comparative assessments discussed previously suggests they might contribute to the subjectivity of performance information. For this reason, I expect that performance benchmarks will serve to induce System 1 processing; we should expect to see public managers exhibit cognitive bias in their interpretations of performance information in the presence of performance benchmarks. I expect benchmarks will drive respondents to provide higher (lower) performance assessments if the raw metric they see is greater (less) than the benchmark.

Decision Justifications

Finally, in line with Seidenfeld's comments on bureaucratic decision making in the rulemaking process (Seidenfeld 1996, Seidenfeld 2001), I expect the need to justify will induce System 2 thinking. The expectation is that having to justify one's thoughts will lead public managers to undergo a more deliberate thought process when considering performance

62

information. Ergo, I expect public managers will be less prone to exhibit psychological biases when they are asked to justify their thought processes.

Experimental Design

I use two experiments to provide evidence to the question of how public sector employees (including managers) use performance information. I collected data across two separate survey instruments. In this section, I will independently discuss the survey instruments used to collect the experimental data as well as the two survey experiments respondents undertook. Experimental vignettes and workflows as well as randomization checks can be found in the appendix.

Pre-Registration

Prior to any research activities, I pre-registered the hypotheses on framing effects and negativity bias with the Evidence in Governance and Politics (EGAP) group under the following ID: 20170501AC. And, I pre-registered the hypotheses for performance benchmarking and the justification requirement with the Evidence in Governance and Politics (EGAP) group under the following ID: 20180425AD.

Experiment 1

Data Collection

Data for this experiment came from a Qualtrics panel collected during May of 2017. I recruited respondents directly through Qualtrics to avoid some of the potential pitfalls of using other online survey platforms (Stritch et al., 2017). Qualtrics screened and provided the respondents for the survey. I provided a stipulation that respondents were managers in their organization. The total sample size is 300, with 150 coming from the private sector and 150 from the public sector. All respondents were initially targeted by a partner of Qualtrics through self-reporting. Then, those responses were screened out to remove misidentified respondents using

63

things like red-herrings to make sure the sample was accurate to qualifications. Qualtrics collected the final data through a partner firm with the ability to prescreen—respondents were asked additional questions at the beginning of the survey to remove individuals whose responses did not match previous identifying responses—and make sure only managers and above could complete the survey. I pre-registered this survey with the Evidence in Governance and Politics (EGAP) group under the following ID: 20170501AC.

Experiment

One of the challenges with experiments is the ability to generalize the findings outside of the experiment. One way I attempted to address this problem was to base the wording of the vignettes off an actual performance management system. In these experiments, the wording for the performance information metrics comes from KCStat, the performance management system dashboard for the city of Kansas City, Missouri. (In fact, I used the same performance prompt from KCStat to serve as the foundation for both experiments.)

Respondents were asked to assume they were serving in a managerial position for a (fictional) city. They were then asked to interpret performance information and provide a performance assessment for the (fictional) city. Respondents saw the following prompt before the experiment:

> 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. While you may see multiple pieces of performance information please only consider the information before you at that time when providing an assessment.

64

This experiment dealt with the city's business development efforts. Respondents were asked to rate the city's performance in this area. All respondents saw the following prompt:

> Imagine that you are the manager of a business development office for a major metropolitan area. Your city just finished compiling the results of a business-related survey that went out to local businesses. In one question business owners were asked whether they were generally satisfied or generally dissatisfied with the city as a place to do business. In addition to the survey results, the city just released its yearly performance metrics. Using the information available to you, use the sliding scale (0-100) to rate the city's performance in regard to business development based upon the following performance information:

Then, each respondent saw one of six randomly assigned pieces of performance information. Individuals were randomized into one of six groups. Two groups only saw a performance metric that indicated the percentage of businesses in the city that were (dis)satisfied with the city as a place of doing business (82% satisfied or 18% dissatisfied)—this provided a baseline for the initial framing. The four other groups saw that same information but then also saw an additional frame on employment figures for the city (8% unemployment or 92% employment). The second frame allowed me to test for differences between those who only saw the negative or positive frame and those who also saw the second frame. This allowed me to test how *strongly* individuals responded to the valanced performance information they saw. Because I wanted to focus on the responses to negativity, the base performance information was equivalent for all respondents (i.e., I did not randomize the numerical value of the original performance information respondents saw). While there is value in randomizing the performance information in framing experiments like this, the primary object was to get at the *strength* of the response to framing (Olsen 2015). This required me to look at the magnitude of the differences of the treatment groups from the control groups. Therefore, based upon the constraints of the sample, I opted to fix the performance information and randomize the

treatment groups. Roughly two-thirds of the respondents saw two valanced performance frames for (un)employment.

Groups 1 (CS: Control, Satisfied) and 2 (CD: Control, Dissatisfied) only saw the percentage of businesses which were satisfied with the city as a place to do business (82% satisfied or 18% dissatisfied). Groups 3 (SU: Satisfied, Unemployment) and 5 (SE: Satisfied, Employment) saw the positive frame but saw information about the unemployment rate (8%) and the employment rate (92%), respectively. Groups 4 (DU: Dissatisfied, Unemployment) and 6 (DE: Dissatisfied, Employment) saw the negative frame and the respective (un)employment information. Respondents then indicated their performance assessment on a sliding scale (0-100). Table 1 shows the randomization of groups as well as the number of respondents that were in each group (from the full sample). For Experiment 1, randomization checks for the following variables can be found in the appendix: age, education, gender, PSM (Public service motivation), PSM_APM (Attraction to policy making), PSM_CPI (Commitment to public interest), PSM_COM (Compassion), PSM_SS (Self-sacrifice), work experience (overall), work experience (private sector), and work experience (public sector) (Perry 1996, Kim 2011).

Table 1 - Randomziation Groups for Experiment 1			
Satisfaction Frame	*Employement Frame*		
	No	Yes	
	Control	Unemployment	Employment
Satisfaction (Positive)	Group 1 (CS)	Group 3 (SU)	Group 5 (SE)
	82% Satisfied	82% Satisfied	82% Satisfied
		8% Unemployment	92% Employment
	N = 50	N = 51	N = 49
Dissatisfaction (Negative)	Group 2 (CD)	Group 4 (DU)	Group 6 (DE)
	18% Dissatisfied	18% Dissatisfied	18% Dissatisfied
		8% Unemployment	92% Employment
	N = 48	N = 54	N = 48

Expectations

My initial expectation is that the way respondents interpret performance data will influence the performance the performance assessment they provide. Under the performance management doctrine, an assumption that performance information is objective would suggest that public managers should not differ in how they respond to performance information (the null hypothesis). Here, the basic hypothesis is that the process of interpretation will lead respondents to assess performance differently, depending on the information they see in their randomly assigned group.

Because respondents in Group 2 saw the negative satisfaction frame, I expect Group 2 will have a lower average performance rating than Group 1. For ease of reference, these two groups will be referred to as the "baseline" scores when comparing against the groups which saw information on employment in the city. Because they saw unemployment—i.e., negative— information, I expect respondents will give Groups 3 and 4 lower average performance rations than their respective "baseline" groups (Groups 1 and 2, respectively). By a similar logic, Groups 5 and 6 should report higher average performance ratings than their respective baselines as well.

The negativity bias suggests the magnitude of the change will be greater for the groups which saw the negative information than those that saw the positive information. Because Group 4 saw two pieces of negative information, it should, on average, have the lowest score of all six groups. I also expect the differences between Group 4 and all the other groups to be the largest when comparing group averages. This is due to the magnitude effects of negative potency. Group 5 should have the highest average value. While I expect it to be higher than its baseline reference group, I want to be transparent that from the outset it was unclear what the expectation

67

should be about the statistical relationship between Group 1 and Group 5. That is, I am not sure if the second positive frame should lead to a mean difference in the performance assessments of the two groups that is statistically significant.

Experiment 2

Data Collection

I used surveys to collect data for this experiment in three phases. Individuals were paid for their participation in all phases. I designed data collection instruments for each phase using Qualtrics. I 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 (Litman et al. 2016).

In the first phase, I ran a short survey that allowed me 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.). I provided representative examples in case the sector type would confuse anyone. In addition to this question, I also asked individuals if they had ever worked in each of the three sectors. Respondents could select "yes" or "no" to specific (individual) questions about each sector. 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. 5342 unique individuals completed this 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

68

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 me 1202 individuals who met these qualifications.

I then sent a second survey to these 1202 respondents. This survey included demographic items and scales for the Big 5 personality items and public service motivation. Yet, the real motivation behind this phase was to try and screen out those who passed the first phase of the survey but were not in the population of interest. Someone could easily provide inconsistent answers over time. Individuals might lie in one of the two phases because they believe they know what researchers are looking for. Or, multiple individuals may use the same MTurk account. I undertook this effort in the hope that I could make a stronger claim about the respondents in the final experiment. I received 773 responses from this wave. Of these, 479 met the qualifications through both waves of the survey. These were the potential respondents who were notified of the opportunity to undertake this survey experiment.

The data from this experiment come from a study which I pre-registered with Evidence in Governance and Politics (EGAP) under the following ID: 20180425AD.

Experiment

Experiment 2 was like Experiment 1. It also used a vignette about business satisfaction rates. Here, I was interested in the role of performance benchmarks and justifications as potential moderators for how individuals process performance information. Here, my contribution is to look at both, separately as well as jointly. Table 2 shows the group assignments across these two treatment conditions. Ultimately, the experiment had a 2x2 factorial design with individuals randomly assigned to one of four groups: (Group 1) control,

69

(Group 2) justification, (Group 3) benchmark, and (Group 4) benchmark and justification. Again, the null hypothesis is that the interpretation of performance metrics should not change based upon randomized group assignment. I expect that it will.

Table 2 - Experiment 2			
Group Assignments by Treatment			
Treatments		*Benchmark*	
		No	Yes
Justify	No	1	3
	Yes	2	4

All individuals saw this prompt:

> For this question, imagine that you are the manager of a business development office for a major metropolitan area. Your city just released its yearly performance metrics and, based on this information, the mayor wants to know how you think the city performed over the course of the last year. For some time, business owners were asked whether they were generally satisfied or generally dissatisfied with the city as a place to do business. In 2011, 57% of business owners indicated they were satisfied with the city as a place to do business. The mayor has tasked your office with improving the business climate in the city.

Individuals in the benchmark groups also saw this sentence at the end of the second paragraph of the prompt:

> The goal has been to increase the percentage of business owners satisfied with doing business in the city to 67%.

All individuals saw a randomly generated performance rating which indicated that between 62% and 72% of business owners were satisfied with the city as a place of doing business over the past year. Randomizing the observed performance metric is an important extension from Experiment 1. It allows me to gain a better understanding of how individuals process performance information across a range of potential performance metrics. This also allows me to use a stationary benchmark. While it would be valuable to consider how public

70

managers respond to moving or even multiple benchmarks (i.e., from more than one political principal), either of these options would add a significant complication to the design and analysis. Nonetheless, I encourage future research to consider these elements.

All individuals were asked to assess the performance of the city as a place of doing business for the past year (based on this data). All individuals had an equal probability of seeing a value that was a) less than the benchmark (5/11), b) equal to the benchmark (1/11), or c) greater than the benchmark (5/11). (Note, the first two groups do not see the benchmark.) This allows me to compare the effects of that specific type of reference point.

Half of respondents also need to justify their performance assessment. All respondents who will justify their responses are told they will have to perform this task before they see the raw performance metric. I expect that individuals who are told they will need to justify their performance assessment will be more likely to engage in System 2 thinking (i.e., more deliberate cognitive processing). Ergo, I expect the notification of justification will help to mitigate or remove the effects of any cognitive biases that arise from framing vis-à-vis performance benchmarks. If these expectation holds, I should see no differences in how Group 2 assesses performance relative to Group 1. Since I expect Group 3 to respond to reference points, I expect that the effects of these reference points on interpretation will be lower for Group 3 than in Group 4.

For Experiment 2, randomization checks for the following variables can be found in the appendix: gender, PSM (Public service motivation), PSM_APM (Attraction to policy making), PSM_CPI (Commitment to public interest), PSM_COM (Compassion), PSM_SS (Self-sacrifice), work experience (overall), and work experience (public sector) (Perry 1996, Kim 2011).

Experiment 1

I begin the discussion of the results of Experiment 1 by providing the average scores by group. You can see these results in Table 3. Even at a first glance, these findings suggest the influence of framing effects and potentially a negativity bias in the assessment of government performance among the sample of professional managers.

Table 3 - Experiment 1: Performance Assessment			
Group	Mean	Std. Dev.	Freq
1	78.90	10.64	50
2	71.17	22.04	48
3	79.08	11.24	51
4	65.61	21.68	54
5	82.71	10.41	49
6	74.94	17.48	48
Total	75.29	17.27	300

For Experiment 1, I conducted a one-way between-subjects ANOVA to compare the effect of the randomly assigned treatment frame on the assessed performance of the city across six conditions: control, positive; control, negative; satisfaction, unemployment; dissatisfaction, unemployment; satisfaction, employment; and dissatisfaction, employment. Results from that test suggested the treatment condition influenced how respondents assessed the city's performance ($F(5, 294) = 7.40$, $p = 0.0000$). Unfortunately, those results also suggested the variance across treatment groups was not equal. For this reason, two other forms of analysis were used to provide a more robust assessment of the relationships between the different groups.

First, concerns about variation necessitated variance ratio tests. These tests provided two important insights. One, there was significant between-group variance in the sample (the primary culprit came in the form of the higher standard deviations in those who saw the

72

dissatisfaction frame). Two, these tests signaled where I would need to add Welch tests to t-tests in comparing across treatment groups.

Table 4 shows the results of the between-group comparisons using t-tests. While Table 4 contains a lot of information, I assessed this was the best way to both effectively and efficiently communicate the results. For this reason, I encourage the reader to follow the description of these results before moving on to look at the table. The table allows for comparisons in the relationship between two respondent groups. Each cell has five pieces of information. The top row of each cell shows the mean group difference in the performance assessment between the two groups. Statistical significant is marked by asterisks, as indicated at the bottom of the table. The second row shows three pieces of information: the *p*-value for the relationship in question, whether I could assume equal variances among the two groups, and whether the t-test was a one- or two-tailed test. Because this experiment looks at framing effects and for the presence of negativity bias, several of the relationships allowed for one-tailed tests. Those that were deemed to not be within the scope of the primary research interest were assessed as two-tailed tests—for example, while the difference between Group 4 and Group 1 is substantively large and statistically significant at the $p < 0.01$ level, I did not set out to assess this relationship.

73

Table 4 - Experiment 1: t-test Comparisons					
Group	1	2	3	4	5
2	-7.73** (0.0157)†^				
3	0.18 (0.4674)‡^	7.91 (0.0290)†^^			
4	-13.29*** (0.0001)†^^	-5.56 (0.1014)‡^	-13.47*** (0.0001)†^		
5	3.81** (0.0373)‡^	11.55*** (0.0016)†^^	3.64** (0.0484)‡^	17.10*** (0.0000)†^^	
6	-3.96 (0.1765)†^^	3.77 (0.1777)‡^	-4.14 (0.1675)†^^	9.33*** (0.0098)‡^	-7.78*** (0.0048)†^
Results from individual t-tests across groups. Values represent the difference in the mean performance assessment between a base group (column) and a comparison group (row). (E.g., the mean of Group 2 is 7.73 less than the mean of Group 1.) P-values in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. † Variance unequal, ‡ Variance not unequal. ^ One-tailed test, ^^ Two-tailed test.					

I will now highlight some of these findings to include a description of the presence and strength of framing effects in the use of public sector performance information among professional managers. I will discuss these results, broadly, before focusing on the statistical significance of these tests. The first way to assess the role of framing effects is to understand that based upon the experimental design, groups 2, 4, and 6 should generally have lower mean performance assessments than Groups 1, 3, and 5, respectively. I can see this in two ways. First, I find that all the negative relationships can be found in those comparisons in which the comparison group is an even number. Second, I can compare the relationship between the following respective groups: 1-2, 3-4, and 5-6. The mean response for each of the groups that saw the dissatisfaction frame was statistically lower than its respective group which saw the satisfaction frame.

I can also get some sense of the role of a negativity bias in these results. The largest difference between two cells of interest comes from Group 3 and Group 4. With a difference of almost 13.5 points, this value was substantively large and statistically significant at the $p < 0.01$ level. I also find that Group 4 has the largest absolute difference in the four relationships for Groups 3-6 and their respective control. But, contrary to my expectation, the difference between Group 2 and Group 4 was not statistically significant. Neither was the relationship between Group 2 and Group 6. The relationship between Group 3 (satisfaction, unemployment) and its baseline category did not meet the initial expectation as it was not less than the control nor was it statistically significant. The only relationship that met expectations on the second framing dimension was the difference between Group 1 and Group 5. It was statistically significant and in the expected direction. Thus, three of the four treatment groups had a mean performance assessment that was in the expected direction compared to its respective baseline but only one of those three groups was statistically distinct from the baseline. I conclude there is marginal evidence in support of a negativity bias in the use of performance information in this sample of professional managers.

Experiment 2

In Experiment 2, a one-way ANOVA suggests no statistically significant relationship between the randomized group assignment and the satisfaction variable (performance metric) individuals observed in this experiment ($F(3(350) = 1.70$, $p = 0.1660$). That is, statistically speaking, group assignment did not influence the observed performance metric. But, a one-way ANOVA on the assessed organizational performance did show evidence for a statistically significant relationship for the randomized group assignment ($F(3(350) = 12.31$, $p = 0.0000$). These first results allow us to take some confidence that the randomized group assignment

influenced how individuals assessed performance through their interpretations of qualitatively similar performance metrics. Table 5 shows results for group means and standard deviations for both the observed performance metric and the assessed performance for each of the four groups.

Table 5 - Experiment 2:					
Business Satisfaction and Organizational Performance					
Group Information		Observed Variable (Embedded Data)		Performance Assessment	
Number	Freq.	Mean	Std. Dev.	Mean	Std. Dev.
1	88	66.41	3.15	66.92	12.81
2	88	67.41	3.38	67.20	11.67
3	92	66.71	3.03	71.27	13.48
4	86	67.16	3.35	76.87	11.19
By group, mean value and standard deviation of the observed satisfaction value and the respondent's assessed performance.					

Seeing evidence that treatment groups influenced the way individuals assessed performance, the next steps will uncover the reasons for those differences. The three variables of interest are 1) whether the respondent saw a performance benchmark (67% satisfaction), 2) whether the respondent was told they would need to justify their response, and 3) the observed satisfaction value (which was also randomized). The first two assume that the treatment group matters, that it influences how respondents process the observed satisfaction metric. Together, these three allow me to tease out why there are differences in how the groups assessed performance.

Table 6 communicates the most important information from Experiment 2. At the treatment level, it compares within both the benchmark and justification treatments. It also compares at the group level, showing which relationships are statistically distinct from one another. All tests of significance in this table were either one-way ANOVA or standard deviation tests, respectively. The importance of the performance benchmark treatment clearly stands out in this table.

76

Broadly, if I look at each treatment class as a binary variable, both treatments appear to provide statistically significant differences between those who saw the treatment and those who did not. In the benchmark treatment, those who did not see the benchmark had a mean performance assessment of 67.06 while those who saw the performance benchmark provided a mean response of 73.98. This is a difference of almost seven points and significant at the $p < 0.01$ level. The difference for those who saw the justify treatment was 2.84 points (No, 69.14; Yes, 71.98) and statistically significant at the $p < 0.05$ level. Moving on to the groups, the mean performance assessments for both Group 3 and Group 4 (benchmark groups) were statistically different than their respective non-benchmark comparisons. But, when I look at the justify treatment, it was only significant within the benchmark treatment. That is, without the benchmark, respondents did not interpret performance any differently depending on whether they would need to justify their performance assessment. This finding suggests some limitations to a need to justify performance assessments as a potential moderator in bureaucratic decision making.

Table 6 - Experiment 2:									
Assessed Performance									
By Treatment Class									
Variable	Group	Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
Benchmark	No	67.06	12.22	0.92	27.22	0.0000	0.92	0.6050	176
	Yes	73.98	12.71	0.95					178
Justify	No	69.14	13.30	0.99	4.31	0.0387	1.15	0.3481	180
	Yes	71.98	12.39	0.94					174
By Treatment Group (focus Benchmark)									
Group		Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
1		66.92	12.81	1.37	4.92	0.0279	0.90	0.6322	88
3		71.27	13.48	1.41					92
2		67.20	11.67	1.24	31.07	0.0000	1.09	0.6981	88
4		76.87	11.19	1.21					86
By Treatment Group (focus Justify)									
Group		Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
1		66.92	12.81	1.37	0.02	0.8780	1.20	0.3865	88
2		67.20	11.67	1.24					88
3		71.27	13.48	1.41	9.02	0.0031	1.45	0.0835	92
4		76.87	11.19	1.21					86
Control vs Both Treatments									
Group		Mean	Std. Dev.	Std. Err.	ANOVA		SD Test		N
					F	Prob > F	f	Pr (F>f)	
1		66.92	12.81	1.37	29.72	0.0000	1.31	0.2124	88
4		76.87	11.19	1.21					86

Moving on to an analysis of how respondents processed specific performance metrics, three separate one-way ANOVA tests suggest a statistically significant difference between the observed satisfaction variable and the performance assessment respondents provided; in the full sample ($F(10(343)) = 6.49$, $p = 0.0000$), in the sub-sample that did not see a performance benchmark ($F(10(165)) = 3.62$, $p = 0.0000$), and in the benchmark sub-sample ($F(10(167)) = 4.19$, $p = 0.0000$). Knowing this, I can dig a little deeper into how respondents processed different performance metrics.

78

Table 7 shows the substantive differences in the means of three different sub-groups based upon the performance metric their respondents saw in relation to the performance benchmark (note, this only pertains to Group 3 and Group 4). These data appear to show that the justification prompt led respondents in Group 4 to provide higher performance assessments compared to Group 3 (this was contrary to the initial expectation that a justification requirement would induce System 2 thinking and thus reduce cognitive bias [i.e., the role of the performance benchmark as a reference point]). For example, respondents in Group 4 provided a mean performance assessment more than 5 points larger than those who saw the same information in Group 3 but were not asked to make a justification of the performance assessment.

Table 7 - Experiment 2: Mean Performance Assessment Relative to Performance Benchmark								
Relation to Benchmark	Full Sample (Groups 3 and 4)			Group 3			Group 4		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Less than	68.85	12	82	67.96	13.56	46	70	9.72	36
Equal to	71.00	14.06	16	65.1	11.59	10	80.83	12.86	6
Greater than	79.83	10.63	80	77.22	11.83	36	81.95	9.12	44
Total	73.98	12.71	178	71.27	13.48	92	76.87	11.19	86

Table 8 shows results for group means and standard deviations for the full sample and two sub-samples: those who did not see the performance benchmark and those who did. This provides more granularity in my comparisons of the role of the performance benchmark. I will discuss a few points that stand out from this table before moving on to another discussion of statistical significance. First, broadly, I have even more detailed evidence that respondents assessed performance differently if they saw a performance benchmark. For example, if I look at those who saw an observed value of 67, those who saw this in the context of the performance benchmark reported a mean performance of more than five points higher than the group that did not see a performance benchmark. Additionally, in the benchmark sub-sample, a performance metric of 67 appears to be an inflection point. But, in the sub-sample that did not see the

79

benchmark the first observed value that produced a statistically different interpretation from other values within this sub-sample was '70'. This may actually suggest respondents were responding to another cognitive bias that can influence the processing of performance information, a left-most digit bias (Olsen 2013). Additionally, the fact that the first significantly different observed value was higher in the performance benchmark again suggests the benchmark inflated the interpretation of performance information.

Table 8 - Experiment 2: Organizational Performance by Observed Satisfaction									
Observed Variable (Embedded Data)	*Performance Assessment (Full Sample)*			*Performance Assessment (No Benchmark Sub-sample)*			*Performance Assessment (Benchmark Sub-sample)*		
Value	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
62	65.60	9.66	35	61.67	7.70	18	69.76	9.97	17
63	65.66	11.26	35	64.70	11.43	20	66.93	11.30	15
64	65.50	12.89	30	59.67	15.27	12	69.39	9.62	18
65	65.60	13.65	40	61.35	11.80	20	69.85	14.32	20
66	67.96	12.96	23	68.45	9.89	11	67.50	15.70	12
67	*68.18*	*14.84*	*34*	*65.67*	*15.45*	*18*	*71.00*	*14.06*	*16*
68	72.76	12.04	37	66.29	10.59	17	78.25	10.54	20
69	72.33	11.77	24	68.54	13.10	13	76.82	8.47	11
70	**78.24**	**10.88**	**29**	**76.18**	**11.98**	**11**	79.50	10.30	18
71	79.43	10.26	28	74.07	6.86	15	**85.62**	**10.22**	**13**
72	76.56	10.88	39	74.00	9.36	21	79.56	12.00	18
Total	70.54	12.92	354	67.06	12.22	176	73.98	12.71	178
Mean performance assessment for respondents who saw the respective business satisfaction value. Table includes means for the whole sample and sub-samples split by whether respondents saw a performance benchmark. Performance benchmark (67) variable shown in italics. For each sub-sample, the first between-group statistical difference is shown in bold.									

Discussion

These findings suggest the importance of research on the question of how public managers interpret and process performance information. Rather than understanding performance metrics as objective, these results suggest differences in how individuals process this performance information—i.e., these data are interpreted subjectively. Despite some very recent findings suggesting otherwise, previous work in performance management suggests

subjectivity arises predominantly from a public manager's role within an organization. But these experiments suggest how information is presented contributes to the information's subjectivity through information process as well. These findings are important and speak to the importance of the design of performance management systems. While political officials may see performance management systems as an accountability tool, these results also suggest it may be easy to overlook the way these systems facilitate the role of cognitive bias to influence how individuals interpret—ostensibly objective—performance information.

Benchmarking Versus Aspirations and Goal-Setting

One area I feel merits discussion, and represents an avenue for future research, is the distinction between performance benchmarking and performance aspirations, or goal-setting. Because here I am talking about "goals", the reader may want to relate benchmarking (as discussed herein) to the literature on aspirations in psychology and goal-setting (theory) in management science. These differ from public sector performance benchmarks in important ways. By benchmarking I mean when an external entity (i.e., a political principal) places, as a comparison, a marker or goal. But, aspirations and goal-setting involve an individual or organization actually going through the process of setting the goal themselves (Lewin et al. 1944). I believe this external benchmarking more closely aligns with the activity that respondents in the experiments undertook because I provided them the performance goals. But, I acknowledge the importance that goal-setting may play in understanding the behavioral foundations of how public managers use performance information (Grimmelikhuijsen et al. 2016). To this extent, I encourage future research to explore a) the role of internally established performance goals in the use (and especially the processing) of performance information and b)

81

potential similarities and differences between these two concepts in the use of performance information.

Conclusion

This paper contributes to a growing literature on the role of cognitive bias in the use of performance information. By studying professional managers and a sample of public sector employees, I add to a literature that seeks to understand how public managers use the information created by performance management systems. Additionally, the use of experimental methods allows me to incorporate an information processing approach. Such an approach is often missing from methods commonly used to study how bureaucrats use performance information (Ammons and Rivenbark 2008, Moynihan et al. 2017).

I believe the research designs speak to challenges inherent in doing this type of research on this specific population. In Experiment 1, I used Qualtrics to recruit a sample of professional managers. This approach offers the benefit of being able to recruit professional managers. But, over the long run there are probably more cost-efficient ways to get at behavioral research questions within the population of interest. In Experiment 2, I tried to navigate the balance between cost and being able to say something of interest about the population of interest. I addressed the issue of cost by collecting the data for this experiment myself, through an online survey. I addressed the issue of external validity by requiring research participants to pass two separate screening instruments, also delivered through online surveys. The screening instruments allow me to take some confidence that the respondents are who they say they are—public sector employees or individuals with significant experience working in the public sector.

The survey experiments of professional managers allow for tests of the role of cognitive bias in bureaucratic decision making and the interpretation of performance information. In

Experiment 1 there is considerable evidence for the role of framing effects in bureaucratic decision making. There is also some evidence about the magnitude of the interpretations professional managers use to interpret information, but it did not align with initial expectations. From the view of psychology, generally, these findings are not surprising. But, they do run counter to some of the general assumptions undergirding performance management systems (Radin 2006, Moynihan 2008).

In Experiment 2, I build off the results from Experiment 1 to ask, knowing that framing can influence how professional managers *interpret* performance information, how do public employees respond to performance benchmarks or the need to justify a performance assessment? As expected, I found strong evidence that performance benchmarks influence the way public sector employees interpret performance information. I found less evidence to support the expectation that justifications will induce a more deliberate decision making process. In fact, these data may suggest the opposite—the need to justify may in fact intensify the role of a performance benchmark as a reference point. While this is an interesting observation, the nature of the experimental design limits my ability to adequately address this question. For this reason, untangling the relationship between performance benchmarks and justifications of decision making provides an avenue for future research.

I believe these findings, especially the mixed findings from Experiment 2, suggest there is value in giving more attention to how bureaucrats process performance information.

References

Ammons, D. N. and W. C. Rivenbark (2008). "Factors influencing the use of performance data to improve municipal services: Evidence from the North Carolina benchmarking project." Public Administration Review 68(2): 304-318.

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.

Askim, J., Å. Johnsen and K.-A. Christophersen (2007). "Factors behind organizational learning from benchmarking: Experiences from Norwegian municipal benchmarking networks." Journal of Public Administration Research and Theory 18(2): 297-320.

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.

Baumeister, R. F., E. Bratslavsky, C. Finkenauer and K. D. Vohs (2001). "Bad is Stronger Than Good." Review of General Psychology 5(4): 323-370.

Behn, R. D. (1995). "The big questions of public management." Public Administration Review: 313-324.

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.

Bertelli, A. M. and L. E. Lynn (2006). Madison's Managers: Public administration and the Constitution. Baltimore, MD, Johns Hopkins University Press.

Bouckaert, G. and W. Van Dooren (2009). Performance Measurement and Management in Public Sector Organizations. Public Management and Governance. T. Bovaird and E. Löffler. New York, Routledge. 2**:** 151-164.

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

Charbonneau, É. and G. G. Van Ryzin (2015). "Benchmarks and citizen judgments of local government performance: Findings from a survey experiment." Public Management Review 17(2): 288-304.

Druckman, J. N. (2001). "Evaluating Framing Effects." Journal of Economic Psychology 22(1): 91-101.

Freeman, R. A. (1954). "How can voters understand?" National Civic Review 43(3): 123-129.

Gaines, B. J., J. H. Kuklinski, P. J. Quirk, B. Peyton and J. Verkuilen (2007). "Same facts, different interpretations: Partisan motivation and opinion on Iraq." Journal of Politics 69(4): 957-974.

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.

Giles, J. (2011). "Social science lines up its biggest challenges." Nature 470: 18-19.

Gormley, W. T. and S. J. Balla (2008). Bureaucracy and Democracy: Accountability and Performance. Washington, DC, CQ Press.

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

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. 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.

James, O. and P. John (2007). "Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments." Journal of Public Administration Research & Theory 17(4): 567-580.

James, O. and G. G. Van Ryzin (2017). "Motivated Reasoning about Public Performance: An Experimental Study of How Citizens Judge the Affordable Care Act." Journal of Public Administration Research and Theory 27(1): 197-209.

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.

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. Nobel prize lecture. 8**:** 449-489.

Kahneman, D. (2011). Thinking, Fast and Slow. New York, Farrar, Straus and Giroux.

Kim, S. (2011). "Testing a revised measure of public service motivation: Reflective versus formative specification." Journal of Public Administration Research and Theory 21(3): 521-546.

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

Kroll, A. (2015). "Explaining the Use of Performance Information by Public Managers: A planned-behavior approach." The American Review of Public Administration: 201-215.

Levin, I. P., G. J. Gaeth, J. Schreiber and M. Lauriola (2002). "A New Look at Framing Effects: Distribution of effect sizes, individual differences, and independence of types of effects." Organizational Behavior and Human Decision Processes 88(1): 411-429.

Levin, I. P., S. L. Schneider and G. J. Gaeth (1998). "All Frames are Not Created Equal: A typology and critical analysis of framing effects." Organizational Behavior and Human Decision Processes 76(2): 149-188.

Lewin, K., T. Dembo, L. Festinger and P. S. Sears (1944). Level of Aspiration. Personality and the Behavior Disorders: A Handbook Based on Experimental and Clincial Research. J. M. Hunt. New York, Ronald Press. 1: 333-378.

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.

Löffler, E. (2001). "Quality awards as a public sector benchmarking concept in OECD member countries: some guidelines for quality award organizers." Public Administration and Development 21(1): 27-40.

Meier, K. J. and J. Bohte (2007). Politics and the Bureaucracy: Policymaking in the Fourth Branch of Government. Belmont, CA, Thomson Wadsworth.

Meier, K. J., N. Favero and L. Zhu (2015). "Performance gaps and managerial decisions: A Bayesian decision theory of managerial action." Journal of Public Administration Research and Theory 25(4): 1221-1246.

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.

87

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., P. A. Nielsen and A. Kroll (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. Experiments in Public Management Research: Challenges and Contributions. O. James, S. R. Jilke and G. G. V. Ryzin. New York, Cambridge University Press**:** 244-269.

Moynihan, D. P., P. A. Nielsen and A. Kroll (2017). Managerial use of performance data by bureaucrats and politicians. Experiments in Public Management Research: Challenges and Contributions, Cambridge University Press.

Moynihan, D. P., P. A. Nielsen and A. Kroll (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. Emperiments in Public Management Research: Challenges and Contributions. O. James, S. R. Jilke and G. G. V. Ryzin. New York, Cambridge University Press**:** 244-269.

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 20(4): 849-866.

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 (2012). "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 (2012). "Setting the table: How transformational leadership fosters performance information use." Journal of Public Administration Research and Theory 22(1): 143-164.

Mussweiler, T. (2003). "Comparison processes in social judgment: Mechanisms and consequences." Psychological Review 110(3): 472-489.

88

Mussweiler, T. and K. Epstude (2009). "Relatively fast! Efficiency advantages of comparative thinking." Journal of Experimental Psychology: General 138(1): 1-21.

Mussweiler, T. and A.-C. Posten (2012). "Relatively certain! Comparative thinking reduces uncertainty." Cognition 122(2): 236-240.

Nielsen, P. A. (2013). Performance Information in Politics and Public Management: Impacts on Decision Making and Performance. Department of Political Science and Government. Aarhus, Denmark, Aarhus University. PhD.

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. (2014). "Learning from Performance Feedback: Performance Information, Aspiration Levels, and Managerial Priorities." Public Administration 92(1): 142-160.

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.

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. (2015). "Citizen (dis) satisfaction: An experimental equivalence framing study." Public Administration Review 75(3): 469-478.

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

Olsen, A. L. (2017). "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." Journal of Public Administration Research and Theory.

Perry, J. L. (1996). "Measuring Public Service Motivation: An assessment of construct reliability and validity." Journal of Public Administration Research and Rheory 6(1): 5-22.

Priem, R. L., B. A. Walters and S. Li (2011). "Decisions, decisions! How judgment policy studies can integrate macro and micro domains in management research." Journal of Management 37(2): 553-580.

Radin, B. (2006). "Challenging the performance movement." Washington DC, Georgetown.

Rozin, P. and E. B. Royzman (2001). "Negativity Bias, Negativity Dominance, and Contagion." Personality and Social Psychology Review 5(4): 296-320.

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.

Seidenfeld, M. (1996). "Demystifying deossification: Rethinking recent proposals to modify judicial review of notice and comment rulemaking." Tex. L. Rev. 75: 483.

Seidenfeld, M. (2001). "The Psychology of Accountability and Political Review of Agency Rules." Duke Law Journal 51(3): 1059-1095.

Simon, H. A. (1939). "The administrator in search of statistics." Public Management 21: 106-109.

Simon, H. A. (1955). "A behavioral model of rational choice." The Quarterly Journal of Economics 69(1): 99-118.

Stritch, J. M., M. J. Pedersen and G. Taggart (2017). "The Opportunities and Limitations of Using Mechanical Turk (Mturk) in Public Administration and Management Scholarship." International Public Management Journal: 1-23.

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

Van de Walle, S. and W. Van Dooren (2011). Introduction: Using Public Sector Performance Information. Performance Information in the Public Sector: How it is Used. W. Van Dooren and S. Van de Walle. New York, Palgrave Macmillan: 1-12.

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.

90

West, W. F. (1995). <u>Controlling the Bureaucracy: The Theory and Practice of Institutional Constraints</u>. Armonk, NY, M.E. Sharpe.

West, W. F. (2004). "Formal procedures, informal processes, accountability, and responsiveness in bureaucratic policy making: An institutional policy analysis." <u>Public Administration Review</u> 64(1): 66-80.

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

Yockey, M. D. and S. M. Kruml (2009). "Everything is relative, but relative to what? Defining and identifying reference points." <u>Journal of Business and Management</u> 15(1): 95.

Appendix

Experiment 1		
Randomization Checks		
Variable	F	Prob > F
Age	0.55	0.7350
Education	1.11	0.3551
Experience (Overall)	0.96	0.4399
Experience (Private Sector)	0.32	0.9025
Experience (Public Sector)	0.26	0.9318
Gender	2.82**	0.0166
PSM	0.24	0.9443
PSM_APM	0.75	0.5859
PSM_COM	0.33	0.8927
PSM_CPI	0.57	0.7206
PSM_SS	0.99	0.4267

This table provides the results of a randomization check for Experiment 1. We provide the mean for each of eleven potential control variables. We ran ANOVA tests on each of the controls as a check on randomization. Of note, only gender did not appear to be adequately randomized at the $p = 0.05$ level. But, ANOVA and regression tests indicated gender did not influence the dependent variable of interest.

Variables: PSM (Public service motivation), PSM_APM (Attraction to policy making), PSM_CPI (Commitment to public interest), PSM_COM (Compassion), PSM_SS (Self-sacrifice). These come from Kim (2011) and Perry (1996).

Experiment 2 Randomization Checks		
Variable	F	Prob > F
Gender	0.01	0.9987
PSM	0.90	0.4423
PSM_APM	1.17	0.3208
PSM_COM	0.94	0.4205
PSM_CPI	1.52	0.2097
PSM_SS	0.50	0.6848
Experience (Overall)	1.73	0.1596
Experience (Public Sector)	2.09	0.1017

This table provides the results of a randomization check for Experiment 2. For each treatment group, we provide the mean afor each of eight potential control variables. We ran ANOVA tests on each of the controls as a check on randomization. Of note, each of the controls appears to pass this check at the $p = 0.05$ level.

Variables: PSM (Public service motivation), PSM_APM (Attraction to policy making), PSM_CPI (Commitment to public interest), PSM_COM (Compassion), PSM_SS (Self-sacrifice); Kim (2011) and Perry (1996).

Experiment 1 - Vignettes and Workflow	
Intro Prompt: (All respondents)	Imagine that you are the manager of a business development office for a major metropolitan area. Your city just finished compiling the results of a business-related survey that went out to local businesses. In one question business owners were asked whether they were generally satisfied or generally dissatisfied with the city as a place to do business. In addition to the survey results, the city just released its yearly performance metrics. Using the information available to you, use the sliding scale (0-100) to rate the city's performance in regard to business development based upon the following performance information:
Respondents were then randomly assigned to one of the following six groups:	
Group 1	82% of businesses were satisfied with the city as a place to do business.
Group 2	18% of businesses were dissatisfied with the city as a place to do business.
Group 3	82% of businesses were satisfied with the city as a place to do business. The current unemployment rate in the city is 8%.
Group 4	18% of businesses were dissatisfied with the city as a place to do business. The current unemployment rate in the city is 8%.
Group 5	82% of businesses were satisfied with the city as a place to do business. The current employment rate in the city is 92%.
Group 6	18% of businesses were dissatisfied with the city as a place to do business. The current employment rate in the city is 92%.

Experiment 2 - Experimental Vignettes and Workflow		
Stages	No Benchmark	Benchmark
Introduction	For this question, imagine that you are the manager of a business development office for a major metropolitan area. Your city just released its yearly performance metrics and, based on this information, the mayor wants to know how you think the city performed over the course of the last year.	
Question Frame (Benchmark randomization)	For some time, business owners were asked whether they were generally satisfied or generally dissatisfied with the city as a place to do business. In 2011, 57% of business owners indicated they were satisfied with the city as a place to do business. The mayor has tasked your office with improving the business climate in the city.	For some time, business owners were asked whether they were generally satisfied or generally dissatisfied with the city as a place to do business. In 2011, 57% of business owners indicated they were satisfied with the city as a place to do business. The mayor has tasked your office with improving the business climate in the city. The goal has been to increase the percentage of business owners satisfied with doing business in the city to 67%.
Justify	No Justify	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. Please only consider the information before you at that time when providing an assessment. After you provide the assessment, on the following screen you will be asked to provide a justification for the performance assessment you just provided. / No Justify
Performance Metric Question (XX% signifies a randomly generated performance metric.)	XX% of business owners were satisfied with the city as a place to do business. Using the information available to you, use the sliding scale (0-100) to rate the city's performance in regard to business development based upon this performance information:	XX% of business owners were satisfied with the city as a place to do business. Remember, the mayor's goal was that 67% of business owners would be satisfied with the city as a place to do business. Using the information available to you, use the sliding scale (0-100) to rate the city's performance in regard to business development based upon this performance information:

How Historical and Social Comparisons Influence Bureaucratic Decision Making: A Survey Experiment of Managers

Abstract: The ways in which managers use performance information is among the most salient topics in the study of public management. Drawing inspiration from several recent studies on the use of performance information by citizens, we adopt a behavioral approach to understand the influence of reference dependence on the interpretation of performance information by managers. Specifically, we run two experiments in a sample of professional managers, which allows us to test whether social and historical comparisons influence how respondents interpret performance information. The results suggest that framing an objective performance metric as poor relative to peer or competitor organizations leads managers to rate overall organizational performance significantly lower than managers in a control group that received the same metric, but no comparative frame. The results support expectations about the importance of social comparisons, particularly in the case of negative deviations from the reference point. The fact that we find no impact of historical comparisons on the interpretation of performance information deviates from recent work on citizen evaluations and suggests differences in the ways in which elites process such information. We conclude with a discussion of the implications of our results for the study of performance management and the behavioral approach to public management more generally.

96

Introduction

In 1995 Robert Behn suggested that one of the big questions facing scholars who study public sector organizations was understanding how "*public managers* use measures of the achievements of public agencies to produce even greater achievements" (Behn, 1995, emphasis added). Yet recently, 20 years after Behn pointed out a major question for researchers in the field, a prominent scholar commented that "we know little about the basic tendency of individuals to incorporate and use performance information" (Moynihan, 2015, p. 33). Kroll (2015a) made a significant contribution to this literature with his review of how managers use performance information, but noted that very few studies incorporated a psychological perspective to understand how public managers *process* performance information (note, as exceptions, Andersen & Moynihan, 2016; Kroll, 2015b).

We seek to contribute to this literature by focusing on the question of relative performance evaluation. Research suggests people do not judge information about performance in absolute terms, but in reference to how an organization performed previously (historical comparison) or how it performed relative to peer institutions (social comparison). To date, however, scholars have focused primarily on the historical and social comparisons made by ordinary citizens and not by organizational managers. Because we know that elites process information differently from average citizens due to expertise and experience, among other factors, these results may not offer an accurate picture of the ways in which public managers use these heuristics when judging performance information.

To address this potential gap in our understanding we adopt a behavioral approach to understand the influence of reference dependence on the interpretation of performance information. Specifically, we run two experiments in a sample of professional managers, which

97

allow us to test whether social and historical comparisons influence how respondents interpret performance information. In line with previous research on citizens (Charbonneau & Van Ryzin, 2015; Olsen, 2017), we provide evidence that social comparisons do more to shape how professional managers evaluate and interpret performance information than historical comparisons. Distinct from studies of citizens, however, we find no evidence that historical comparisons have a significant influence on the way in which professional managers interpret performance information. These results provide potentially important insights regarding the design of systems to provide performance information to managers and suggest ways in which performance information should be framed for different audiences. Moreover, the discrepancies between our results and those of recent research suggests that findings from studies of citizens may not tell us everything we need to know about the use of performance information by managers.

The remainder of the paper proceeds in four parts. First, we review the scholarly literature on the use of performance information. Next, we discuss the advantages of an experimental approach and explain our experimental research design. We then describe two experiments and present their results. Finally, we conclude with a discussion and some practical and scholarly implications of this research.

Performance Information
Bureaucrats and the "Use" of Performance Information
From both academic and practical perspectives, one of the compelling aspects of performance management systems is that they generate, at least ostensibly, an objective assessment of how well an organization is doing its job (Nielsen, 2013). And yet, *who* or *what organization* assesses performance metrics potentially influences how the information is

98

interpreted (Moynihan, 2008). The subjectivity of performance information use is one of the key reasons why significant questions remain regarding the ways in which "objective" performance metrics actually influences organizational outcomes.

While scholars study the use of performance information across three groups of "end users"—citizens, managers, and politicians (Van de Walle & Van Dooren, 2011), public managers might represent the main or primary users of performance information. At the very least, they are a key target of the information that nearly ubiquitous performance measurement systems produce. Despite their importance, there is still much to learn about how public managers use performance information (Kroll, 2014, 2015a), which is why we make them the focus of this study.

The literature to date on this subject has clearly established the subjectivity of performance information use by government officials. In an early study, 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 (ibid, p. 588).

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. In other work, Moynihan (2008) suggests that performance information is selected and presented in order to persuade others. For Moynihan, performance information become subjective because actors add their own interpretation to the data. He argues this plays an important role in how we should understand performance information use.

99

As we discuss later, we extend this logic of the importance of interpretation as a part of a cognitive process of performance information use.

A recent systematic literature review offers another set of factors that help us understand why, given their discretion in doing so, bureaucrats sometimes use performance information (Kroll, 2015a). Over the last 15 years in "a highly relevant and fast-growing research area" (ibid., 460), research consistently shows six factors commonly drive the use of performance information among bureaucrats: measurement system maturity (for examples see Berman & Wang, 2000; Ho, 2006; Taylor, 2009), stakeholder involvement (for examples see Bourdeaux & Chikoto, 2008; Ho, 2006; Moynihan & Pandey, 2010), leadership support (for examples see Moynihan & Ingraham, 2004; Moynihan & Lavertu, 2012; Yang & Hsieh, 2007), support capacity (for examples see Berman & Wang, 2000; de Lancer-Julnes & Holzer, 2001; Moynihan & Hawes, 2012), innovative culture (for examples see Moynihan, 2005; Moynihan & Pandey, 2010; Moynihan, Pandey, & Wright, 2012b), and goal clarity (for examples see Moynihan & Landuyt, 2009; Moynihan, Pandey, & Wright, 2012a; Moynihan et al., 2012b).

For our purposes, the most important takeaway from this review of previous work on performance information use is that a majority of studies of how bureaucrats use performance information look to organizational behavior, organizational theory, or a combination of these approaches to explain the phenomenon. In other words, they focus mainly on the ways in which the organizational context of performance measurement systems influences the use of performance information.

Alternatively, recent scholarship suggests that *individual behavioral* factors might produce systematic variation in the use of performance information among bureaucrats (Kroll, 2015a; Moynihan, Nielsen, & Kroll, 2017). For example, several studies over the past decade

100

suggest psychology may play a role in helping us understand how bureaucrats use performance information (Andersen & Moynihan, 2016; Kroll, 2015b; Moynihan, 2008, 2015; Nielsen, 2013; Salge, 2011). In addition to these, some very recent pieces also demonstrate the utility of incorporating an individual-level behavioral approach to examine the use of performance information.

A significant majority of these studies look at how citizens respond to performance information (Andersen & Hjortskov, 2015; Bækgaard & Serritzlew, 2016; Barrows, Henderson, Peterson, & West, 2016; Hvidman & Andersen, 2016; Olsen, 2013, 2015a, 2017), but other work has looked at how politicians (George, Desmidt, Nielsen, & Bækgaard, 2016; Nielsen & Bækgaard, 2015; Nielsen & Moynihan, 2017; Olsen, 2014) use performance information as well. Importantly, some of this work has taken place in the context of education, which is the same service area we focus on in our experiments. These studies have demonstrated that benchmarking plays a

There are three important takeaways from these studies. First, experimental methods are a useful approach to develop our understanding of performance information use across a variety of political actors (Anderson & Edwards, 2015; Bouwman & Grimmelikhuijsen, 2016; James, Jilke, & Ryzin, 2017; Jilke, Van de Walle, & Kim, 2016). Second, when it comes to performance metrics, these studies suggest that, depending upon the circumstances, individuals exhibit various cognitive biases and utilize several heuristics when responding to performance information. Evidence for cognitive bias in the use of performance information by various actors supports the value of taking an information processing approach. Finally, an examination of the recent literature reveals that a relatively limited amount of work has taken a behavioral approach to understanding the use of performance information by public managers. Because these people

101

are a primary target of this information, understanding how cognitive biases influence their assessment of performance metrics is, we believe, an important contribution to the literature.

Reference Points and Information Processing

While there are obviously a number of cognitive biases that may influence assessments of performance information, we focus on the use of reference points and particularly negative deviations from accepted referents in this study. The idea of reference points in the interpretation of information is nothing new, as we demonstrate below, but to date this concept has not been used to understand how public managers interpret performance information.

Psychologists have long understood that human judgment is fundamentally comparative in nature (Mussweiler, 2003). More specifically, we know that individuals make temporal (Albert, 1977) and social (Festinger, 1954) comparisons when evaluating abilities, information, and opinions. Reference points influence decisions because human perception is "*reference-dependent*" (Kahneman, 2002, 459, emphasis in original). These references serve as "stimuli of known attributes that act as standards against which other categorically similar stimuli of unknown attributes are compared in order to gain information" (Yockey & Kruml, 2009, 97). Consistent with the idea that limitations in human processing constrain our ability to accept, hold, and process information (Freeman, 1954; Simon, 1955), reference points also serve as cognitive heuristics in making evaluative judgments about information (Mussweiler & Epstude, 2009; Mussweiler & Posten, 2012).

Historical and Social Reference Points, Negativity Bias, and the Interpretation of Performance Information by Professional Managers

For a variety of reasons, it is reasonable to assume that the use of reference points also influences the interpretation of *performance* information. Indeed, Herbert Simon argued that "the

only sound basis for decisions about numbers is numerical factual information about past experiences or the experiences of others—nothing more nor less than comparative statistics" (Simon, 1939, 106). Reference points facilitate those comparisons, but there are numerous comparative reference points that public managers might use as references. Consistent with Recent work on citizen evaluations of public sector performance we focus on historical and social performance comparisons as points of reference in this paper (Charbonneau and Van Ryzin, 2015; Olsen, 2015b). Historical reference points allow comparison of the performance of an organization to the previous performance of the same organization. In other words, past performance provides a salient status-quo against which individuals can easily assess change and decide if current performance is acceptable. This type of historical reference point is common in performance measurement systems, such as No Child Left Behind, which judges schools on progress relative to the previous year, though it is important to reiterate that our primary interests is in the degree to which such historical comparisons influence the assessment of performance information. Following studies of citizen interpretation of performance information, we hypothesize that providing information about better (worse) past performance of a public-sector organization will lower (raise) a bureaucrat's assessment of the organization's current performance. Olsen (2017),

Social reference points provide another frame in which individuals can compare performance. With a social reference point, individuals compare the performance of their organization against the performance of other, comparable organizations at the same point in time. Often, we might think of these as peer organizations, competitors, or simply organizations in a similar geographic region. The power of social comparisons has received a great deal of empirical support. Festinger (1954) suggested that "people evaluate their opinions and abilities

103

by comparison respectively with the opinions and abilities of others" (118). Charbonneau and Van Ryzin (2015) and Olsen (2017) demonstrate that social comparisons can influence the ratings an individual gives to a public organization. Interestingly, they also found that individuals seem to give more weight to social rather than historical comparisons when using performance information. We expect that providing information about better (worse) performance relative to other, comparable, organizations will lower (raise) a manager's assessment of an organization's current performance.

In addition to our expectations about how public managers will interpret performance information when it is presented as a historical or social comparisons, we are interested in whether the *direction* of the deviation from the reference point matters. Scholars have long recognized that human beings tend to respond more strongly to negative information than to comparable information that is framed in a positive way (Baumeister et al., 2001; Ito et al. 1998; Skowronski and Carlston, 1989). This is because negative information is generally more salient and more potent than positive information (Rozin and Royzman, 2001), thus drawing more cognitive processing.

These expectations have received considerable support in recent research on both citizens' and politicians use of performance information (Boyne et al., 2009; Charbonneau and Bellavance, 2014; Hood, 2007; James, 2011a; James and John, 2006; James and Moseley, 2014; Olsen, 2015; (Nielsen and Bækgaard, 2015; Nielsen and Moynihan 2017). We expect that professional managers will also exhibit a negativity bias and respond more strongly to lower performance relative to an accepted reference point than to an equivalent, but positive deviation.

Experimental Design and Method
Participants

Data for the two experiments come from a Qualtrics panel collected during May of 2017. We recruited respondents directly through Qualtrics to avoid some of the potential pitfalls of using other online survey platforms (Stritch et al., 2017). Qualtrics screened and provided the respondents for the survey. We provided a stipulation that respondents were managers in their organization. The total sample size is 300, with 150 coming from the private-sector and 150 from the public-sector. All respondents were initially targeted by a partner of Qualtrics through self-reporting. Then, those responses were screen out to remove misidentified respondents using things like red-herrings to make sure the sample is accurate. Qualtrics collected our final data through a partner firm with the ability to prescreen—respondents were asked additional questions at the beginning of the survey to remove individuals whose responses did not match previous identifying responses—and make sure only managers and above were allowed to complete the survey.

Our sample includes managers from both sectors because another experiment in the survey required this sector breakdown. But here we will primarily focus on the public-sector managers. We pre-registered the survey with the Evidence in Governance and Politics (EGAP) group under the following ID: 20170501AC.

Olsen (2017) ran a set of experiments like ours involving Danish citizens. He had 3,443 respondents for both of his surveys. Our focus on managers allows us to extend this research using, as respondents, a group of individuals who are likely more accustomed than regular citizens to seeing, thinking about, and using performance information in their decision making. Our sample consists of 300 respondents, evenly *divided* between public and private managers. Below, in Table 1, we provide the descriptive statistics for the combined sample. With an average age of nearly 46 years, and an average of more than 25 years in the workforce, this

105

group of respondents clearly has a significant amount of life and work experience. It is also an educated sample. Roughly 29% have a bachelor's degree and 64% of the sample have at least a bachelor's degree. We can also see that just under half are female.

Table 1 - Descriptive Statistics					
Full Sample					
Variable	Obs	Mean	Std. Dev.	Min	Max
Age	300	45.94	12.37	19	74
Education	300	5.16	2.11	1	9
Female	297	0.49	0.50	0	1
Public-sector Manager	300	0.50	0.50	0	1
Years in Private-sector	217	14.87	12.59	0.5	51
Years in Public-sector	186	17.26	11.41	1	45
Years in Workforce	300	25.25	12.52	1	58
Private sector Managers					
Variable	Obs	Mean	Std. Dev.	Min	Max
Age	150	45.23	121.84	19	74
Education	150	4.76	2.01	1	9
Female	149	0.51	0.50	0	1
Years in Private-sector	126	19.03	13.73	1	51
Years in Public-sector	43	12.52	11.37	1	40
Years in Workforce	150	24.48	12.96	1	58
Public sector Managers					
Variable	Obs	Mean	Std. Dev.	Min	Max
Age	150	46.66	11.89	19	74
Education	150	5.57	2.14	1	9
Female	148	0.47	0.50	0	1
Years in Private-sector	91	9.12	7.81	0.5	37.75
Years in Public-sector	143	18.68	11.08	1	45
Years in Workforce	150	26.02	12.06	1	55

Design of the Experiments

In the design of our experiments, we chose to use the substantive policy area of education—specifically, passing rates for standardized test scores—and modeled the performance information on real test score data to make the experiment more plausible and generalizable to real-world decision making. Specifically, we utilized publicly available data

106

from public schools in the state of Indiana.[6] Data for 2011 and 2012 suggested an average change in the pass rate for English and Math exams to be roughly 2 percent. We averaged the pass rates for schools in the state for both English and Math standardized exams. Since our experiments also include historical comparisons, we also compared school test results across time. Doing this led us to our chosen performance metrics—77% passed the English exam and 79% passed the Math exam—as well as the historical comparison data (2% change from last year).

Due to certain constraints of our research project—namely, recruiting a sample of 300 professional managers—we wanted to address two concerns. The first concern involved the potential for the first experiment to influence the way respondents approached the second experiment. We addressed this by separating the two experiments. Specifically in this regard, respondents saw *Experiment I* near the beginning of the survey and *Experiment II* near the end of the survey. The average response time for the survey across all 300 respondents was 24 minutes and 22 seconds. This meant there was a significant amount of time between these two experiments. We believe this was an adequate way to address any concerns about the first experiment influencing the results of the second. Another notable difference between the two experiments was that in providing the social comparisons in *Experiment II,* we only indicated if the school was in the top- or bottom-half of local schools. That is, we did not include a rank (e.g., 3rd out of 10). This omission was deliberate and an attempt to help respondents not conflate the two experiments despite their similarities. It also allowed us to address a potential concern about causal inference. Specifically, if a respondent sees a prompt that reads "top half", we wouldn't know if she interprets this as first out of three or 49th out of 100.

---

[6] https://www.doe.in.gov/assessment/istep-results

The focus on the educational context in our experiments builds on recent work that takes a behavioral approach to *citizen's* assessments of public organizations. Specifically, builds on a set of studies that explore the influence of performance targets and relative performance information on those assessments (Barrows et al., 2016; Charbonneau and Van Ryzin, 2015).[7] These similarities will allow us to compare the use of performance information by public managers with those of citizens in a comparable service delivery area, which we believe constitutes another contribution for the study.

Before moving on, it is important to note that while we do focus on managers rather than citizens as subjects, our design does not allow us to test the influence of reference effects on their assessments of performance information in their own organizations. Despite this, we believe that the use of managers offers unique insights into the role of hisortical and social comparisons in performance information use for a number of reasons. First, whether being asked to consider their own organization or not, professional managers are likely far more familiar with the types and uses of performance information than are citizens. Borrowing from work on political psychology, we can therefore consider professional managers as "sophisticates" because of their experience with performance information. Research suggests that sophisticates process information differently and more effectively than non-sophisticates (Gaines et al., 2007). They make different decisions (Luskin, 1987; Mintz et al., 2006) and are better able to connect new information to existing knowledge and to relevant decisions (Jerit et al., 2006). Given these differences, it is reasonable to expect that cognitive biases may influence the assessments of

---

[7] It is important to note that there is also a large literature on the influence of performance information on citizen satisfaction with schools (Jacobsen and Saultz, 2016; Jacobsen et al., 2013; Jacobsen et al., 2014; Wilson and Piebalga, 2008). While obviously important, the purpose and findings from these studies are distinct from the purpose of this study.

performance information by managers, who are more sophisticated in the use of such information, in systematically different ways that they influence the assessments of citizens.

Experiment I

In Experiment I we asked respondents to rate the performance of an unnamed high school (High School A) using performance data from a standardized English exam. The goal of *Experiment I* was to observe the assessed performance when both historical and social comparison information were presented together. We felt this would be a suitable way to design the experiment for two reasons. First, in a realistic organizational decision-making environment (i.e., a non-experimental setting), managers might have a sense of their organization's performance as well as the performance of peer and competitor organizations. Second, by including both comparison types in the same experimental frame we might be able to get some sense of the strength of the positive and negative versions of each comparison. Of course, we were also able to compare performance assessments against the control group as well.

In this experiment, individuals saw a raw performance metric which stated that 77% of students at High School A passed the English exam. Respondents were then randomly assigned to one of five groups. The control group saw only the raw performance metric. The other groups saw four combinations of historical and social comparisons. The historical comparison prompts said that the performance was indicative of a 2% increase or decrease in the rate of students who passed the standardized English exam. The social comparison indicated that based upon the pass rate that the school ranked third or seventh out of ten comparable local schools. For the social comparison prompt, individuals were told if this was in the top- or bottom-half of local schools, respectively. Individuals were then asked to rate the performance of the school using a 0-100 sliding scale.

109

As an example, someone in the group that saw prompts indicating increases for both the historical and social comparisons saw the following prompt:

> *"English Exam: 77% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10). Assuming this is the only information available to you, use the sliding scale (0-100) to <u>assess the overall performance of HIGH SCHOOL A over the last year:"</u>

Respondents would rate the performance of the school with the sliding scale. Experimental vignettes for both experiments can be found in the Appendix.

Experiment II

In *Experiment II* we asked respondents to rate the performance of an unnamed high school (High School B) using performance data from a standardized Math exam. In *Experiment II* we wanted to look at the comparisons individually so that we could get a sense of the strength of the comparisons by themselves in the assessment of performance data. Again, respondents were randomly assigned to one of five groups.

To create some generalizability across the two experiments, we used similar comparisons from the first experiment. Individuals saw a raw performance metric that stated that 79% of students at this high school passed the Math exam. As before, the control group saw only the raw performance metric. The other groups saw one of four possible historical and social comparisons. That is, groups 2-5 only saw one of the following: 2% increase from last year, 2% decrease from last year, top-half of comparable local schools, or bottom-half of comparable local schools. Again, respondents were asked to rate the performance of the school on a 101-point sliding scale.

110

Our primary analytic strategy includes ANOVA tests, mean difference tables, figures with means and 95% confidence intervals, and Bonferroni cross group comparison tests, which is consistent with the approach recommended in recent behavioral public administration work (James et al., 2017). As a robustness check, we also include regression models including pretreatment covariates in the Appendix. As Table 2 suggests, however, randomization of subjects across treatment and control groups was adequate. The Qualtrics sample of respondents had both private-sector and public-sector managers. We present results for the full sample with some extra discussion about the managers employed in the public sector.

Table 2 - Randomization Checks											
Experiment I - English Exam - Randomization Check, by Treatment Group											
Treatment	Control		(H) Increase,		(H) Increase,		(H) Decrease, (S)		(H) Decrease, (S)		ANOVA	
Controls	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	F-statistic	Prob > F
Age	45.12	13.44	45.98	12.11	44.74	12.35	47.18	11.74	46.79	12.42	0.44	0.78
Education	4.88	1.97	4.96	2.30	5.32	2.11	5.37	2.17	5.25	2.04	0.65	0.63
Female	0.46	0.50	0.46	0.50	0.57	0.50	0.44	0.50	0.51	0.50	0.72	0.58
Yrs for profit	16.61	14.90	15.60	13.40	14.02	11.10	13.74	12.91	14.48	10.55	0.40	0.81
Yrs public	16.40	12.95	19.32	11.11	13.78	10.45	20.15	11.13	17.37	11.02	2.01	0.10
Yrs workforce	24.34	13.35	25.50	12.04	24.03	11.75	26.69	13.09	25.75	12.55	0.46	0.76
This table provides the results of a randomization check for Experiment 1. For each treatment group we provide the mean and standard deviation for each of six potential control variables. We ran ANOVA tests on each of the controls as a check on randomization. Of note, each of the controls appears to pass this check at the $p = 0.05$ level.											
Experiment II - Math Exam - Randomization Check, by Treatment Group											
Treatment	Control		(H) Increase		(H) Decrease		(S) Top Half		(S) Bottom Half		ANOVA	
Controls	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	F-statistic	Prob > F
Age	44.90	12.85	47.74	12.59	47.10	12.44	44.29	13.09	45.64	10.83	0.83	0.51
Education	5.03	2.11	4.77	2.41	5.36	2.22	5.55	1.86	5.14	1.90	1.21	0.31
Female	0.50	0.50	0.43	0.50	0.52	0.50	0.48	0.50	0.52	0.50	0.34	0.85
Yrs for profit	17.85	15.08	14.44	12.34	16.64	12.34	12.84	12.53	12.68	10.01	1.45	0.22
Yrs public	14.94	11.22	19.47	10.55	18.22	12.81	17.27	12.80	16.46	9.83	0.96	0.43
Yrs workforce	24.45	13.72	27.69	13.02	26.57	11.86	22.28	12.36	25.15	11.07	1.65	0.16
This table provides the results of a randomization check for Experiment II. For each treatment group we provide the mean and standard deviation for each of six potential control variables. We ran ANOVA tests on each of the controls as a check on randomization. Of note, each of the controls appears to pass this check at the p = 0.05 level.											

Experiment I

For Experiment I, we conducted a one-way between-subjects ANOVA to compare the effect of historical and social comparison performance information on assessed performance of a high school in five conditions: control; increase (historical), upper half (social); increase (historical), bottom half (social); decrease (social), upper half (social); and decrease (historical),

bottom half (social). Table 3 provides the one-way ANOVA results for the full sample of respondents, as well as for both the private and public sector respondents separately. As the table suggests, there was a statistically significant effect for the independent variable on the dependent variable at the $p < .05$ level for the five conditions in the full sample as well as within each of the sub-samples: private and public sector managers.

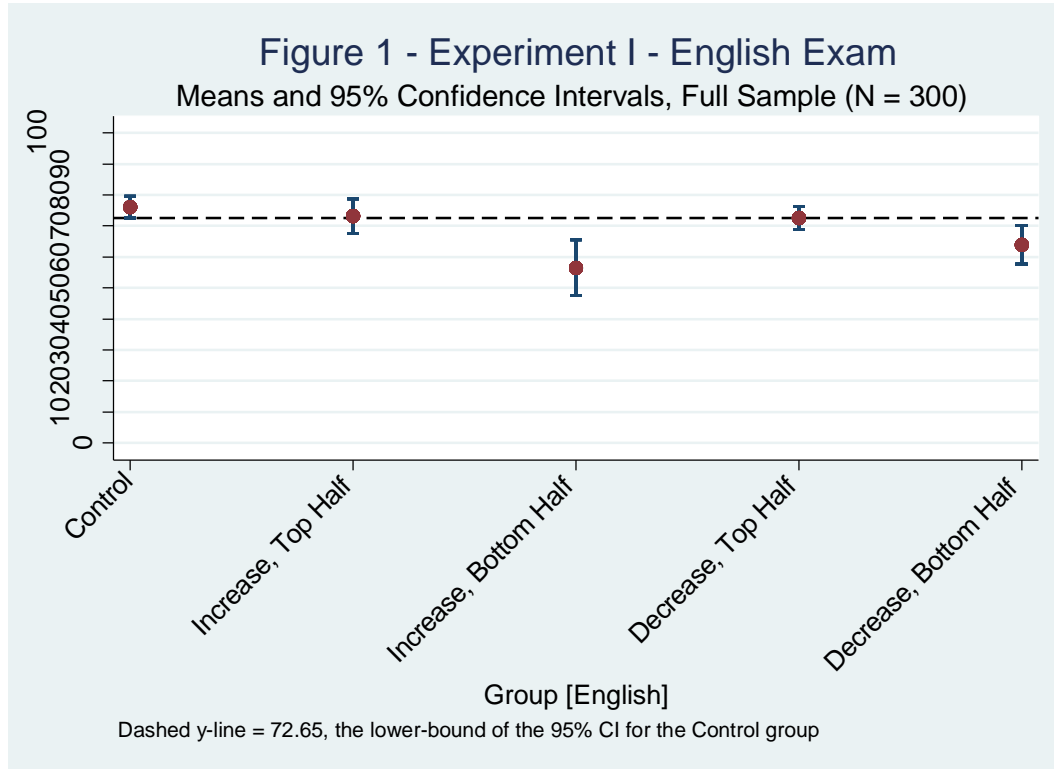| Table 3 - Experiment 1 - English Exam				
Sample	df_between	df_within	F ratio	*p-value*
Full	4	295	10.68	0.0000
Private	4	145	4.57	0.0017
Public	4	145	7.36	0.0000

While the statistically significant findings for the treatments in the one-way ANOVA are intriguing, for the purposes of hypothesis testing we are most interested in which specific groups were statistically distinct from one another. We will present the evidence for these differences in several ways. First, Table 4 presents the means of assessed performance of "High School A" for each experimental vignette across all three samples. The different treatments (upper half (social), increase (historical); upper half (social), decrease (historical); etc.) are presented in the rows and means that are statistically distinct from the control vignette are marked with an asterisk. The results suggest that respondents reacted most clearly to the negative social comparison. The assessment of performance information by that group was significantly lower than the control group, regardless of whether the historical performance was increasing or decreasing. It is important to remember here, that all groups were given the *same* objective information about the school's performance. We do not see significant differences from the control group in the assessments of performance information in either of the groups exposed to positive social comparison, regardless of the presentation of historical information (increasing or decreasing)

112

they received. It is also worthwhile to note that none of the differences between the two sub-sample groups (public vs. private) were statistically different from one another.

Vignette	Sample		
	Public Sector	*Private Sector*	*Full Sample*
Control	76.15	73.12	74.46
2% Inc (H), Upper (S)	73.21	76.6	74.4
2% Inc (H), Lower (S)	56.52*	62.97*	60.29*
2% Dec (H), Upper (S)	72.56	74.57	73.53
2% Dec (H), Lower (S)	63.86*	61.34*	62.58*

Table 4 - Experiment I
Mean Performance Assessment

Values represent the mean of the assessed performance of "High School A" by treatment group. An * indicates those groups which are statistically distinct from the "Control" vignette at the $p < 0.05$ level.

For ease of interpretation, these results are presented graphically in Figure 1. The dashed line represents the lower bound of the 95% confidence interval for the assessed performance in the control group. The 95% confidence intervals for both groups that saw the "bottom half" social comparison fall below this line, suggesting that this treatment causes respondents in these groups to assess the performance of the school more negatively than the control group.

113

Figure 1 - Experiment I - English Exam

Means and 95% Confidence Intervals, Full Sample (N = 300)

Group [English]

Dashed y-line = 72.65, the lower-bound of the 95% CI for the Control group

Because the independent variable was categorical, we ran three different post-hoc analyses including the Bonferroni, the Scheffe, and the Sidak. The results across all three tests are similar so, in the interest of brevity, we present only the Bonferroni analysis in Table 5. The analysis suggests that the differences across groups are substantively meaningful. Across the full sample, the group which saw the historical increase and the negative social comparison had a mean response that was more than 14 points lower than the mean of the control group. This is a difference of 3.19 standard deviations. And, the group which saw the historical decrease and the negative social comparison had a mean response that was almost 12 points lower than the mean of the control group. This is a difference of 2.82 standard deviations. It is also worth noting that the results for the statistically significant differences between the respondents in the control

114

group and respondents in the two groups which saw the negative social comparison were robust across all three sample groups: full sample, private sector, and public sector. [8]

Table 5 - Experiment I - English Exam - Bonferroni by Treatment Group Comparisons				
Treatment Groups	*Control*	*Increase (H), Upper Half (S)*	*Increase (H), Bottom Half (S)*	*Decrease (H), Upper Half (S)*
Increase (H), Upper Half (S)	-.054 1.000			
Increase (H), Bottom Half (S)	-14.165* 0.000	-14.112* 0.000		
Decrease (H), Upper Half (S)	-.925 1.000	-.871 1.000	13.24* 0.000	
Decrease (H), Bottom Half (S)	-11.878* 0.002	-11.825* 0.002	2.287 1.000	-10.953* 0.004
Full sample comparisons: the raw numbers in each comparison represent the difference between the column group from the row group. "H" and "S" represent historical and social treatment conditions, respectively. Numbers marked with an * represent comparisons in which the difference between the groups was statistically significant at the $p < 0.05$ level.				

Experiment II

As a reminder, in this case we present respondents with *either* an historical or social comparison, in order to allow for an assessment of their independent influence on the interpretation of performance information. We again begin the analysis with a one-way between-subjects ANOVA to compare the effect of historical or social comparison performance information on the assessed performance of a high school in five conditions: control; increase (historical); decrease (historical); upper half (social); and bottom half (social). The results, presented in Table 6, suggest a statistically significant effect ($p < 0.05$) for the treatment in the full sample as well as within each of the sub-samples (private and public sector managers).

---

[8] A one-way ANOVA suggests that this sample suffers from unequal variance, so as a final robustness so we also run a Kruskal-Wallis test, which is a nonparametric alternative (Hamilton, 2008, p. 165). The findings are essentially identical to those reported in Table 3.
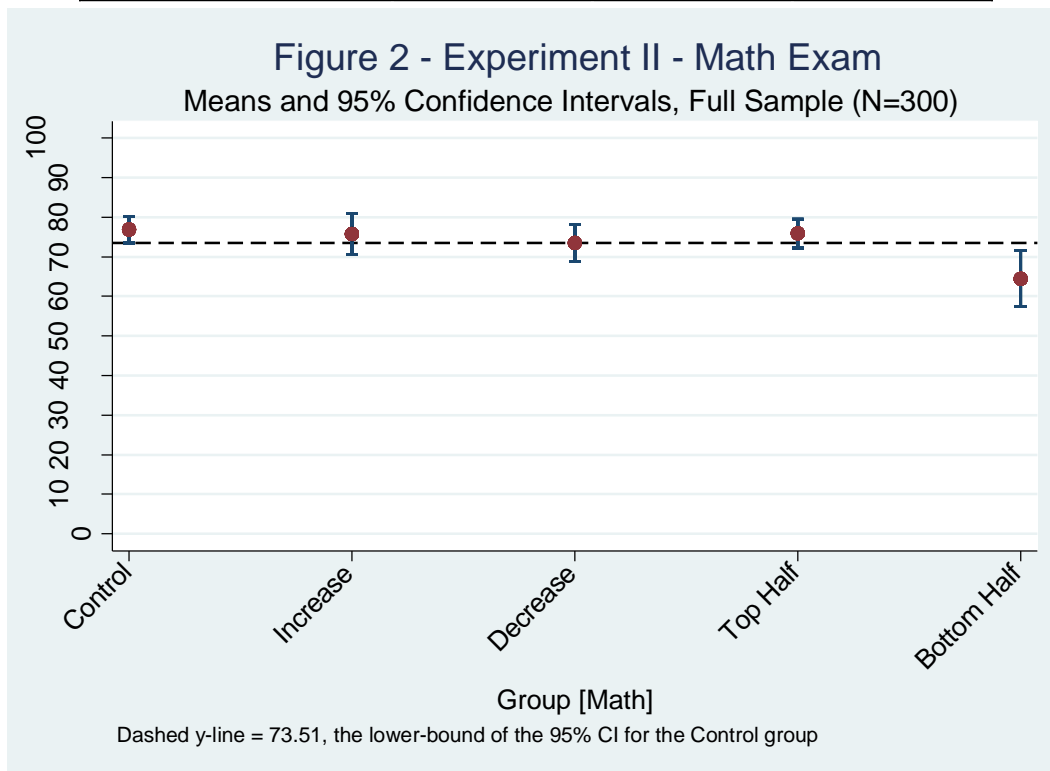
115

Table 6 - Experiment 2 - Math Exam				
Sample	df_between	df_within	F ratio	*p-value*
Full	4	295	10.48	0.0000
Private	4	145	6.77	0.0001
Public	4	145	4.03	0.0039

Again, however, our real interest is in the degree to which different reference points influence the assessment of performance information and, so, we present the mean response by survey vignette in Table 7. The different groups are presented in the rows of the table and significant differences are marked by an asterisk. The findings suggest that respondents in the "bottom half" social comparison provided performance assessments that were significantly lower ($p<0.05$) than those provided by the control group. This was true for the full sample and held within each of the two sub-samples as well. No other groups provided responses which were meaningfully different from that of the control group. For ease of interpretation, we also present the results of the mean comparison graphically in Figure 2. The dashed y line represents the lower bound of the 95% confidence interval for the assessed performance in the control group. The only group for which the 95% confidence intervals fall below this line is the "bottom half" social comparison.

This result reinforces the findings from the first experiment that negative social comparisons significantly influence how managers assess performance information. At the same time, the results of this experiment do not support our first hypothesis that managers will respond to historical reference points.

www.manaraa.com

Table 7 - Experiment 2 Mean Performance Assessment			
Vignette	Sample		
	Public Sector	*Private Sector*	*Full Sample*
Control	76.82	75.34	76.14
2% Inc (H)	75.76	72.29	74.42
2% Decrease (H)	73.44	74.23	73.86
Upper Half (S)	75.84	80.03	78.22
Bottom Half (S)	64.5*	62.36*	63.31*
Values represent the mean of the assessed performance of "High School A" by treatment group. An * indicates those groups which are statistically distinct from the "Control" vignette at the p < 0.05 level.			

Figure 2 - Experiment II - Math Exam
Means and 95% Confidence Intervals, Full Sample (N=300)

Group [Math]

Dashed y-line = 73.51, the lower-bound of the 95% CI for the Control group

To confirm the robustness of the result, we again present a Bonferroni comparison of the control and treatment groups (Table 8). When we consider our full sample, the bottom half social comparison is statistically distinct from the control group as well as all the other treatment groups in this comparison. None of the other treatment groups are statistically different from the control group. In the full sample these comparison results were robust across the Scheffe and Sidak comparisons. Additionally, across the full sample, the group which saw the "bottom half"

117

social comparison had a mean response that was almost 13 points lower than the mean of the control group. This is a difference of 2.59 standard deviations, which again represents a substantively meaningful impact.[9]

Treatment Groups	Control	Increase (H)	Decrease (H)	Upper Half (S)
Increase (H)	-1.724 1.000			
Decrease (H)	-2.281 1.000	-.557 1.000		
Upper Half (S)	2.081 1.000	3.805 1.000	4.362 0.882	
Bottom Half (S)	-12.838* 0.000	-11.114* 0.000	-10.557* 0.000	-14.919* 0.000

Table 8 - Experiment 2 - Math Exam - Bonferroni by Treatment Group Comparison

Full sample comparisons: the raw numbers in each comparison represent the difference between the column group from the row group. "H" and "S" represent historical and social treatment conditions, respectively. Numbers marked with an * represent comparisons in which the difference between the groups was statistically significant at the p < 0.05 level.

Discussion

Motivated by a growing research program in public administration on the use of performance information, we design a set of experiments to understand how reference points, or the comparison of metrics to pre-established benchmarks of acceptability, influence that process. Specifically, we explore whether comparisons of performance information against previous performance (historical comparison) or the performance of peer institutions (social comparison) influences the interpretation of that information. We build on previous work in the public

---

[9] As we did in *Experiment I*, we ran a Kruskal-Wallis test because of our concerns about unequal variance in the dependent variable (the assessed performance rating of High School B given the pass rate on the Math exam). Again, these tests indicated a difference of means, with the clear distinction coming from the lower half social comparison.

118

administration literature, which focuses on the use of historical and social comparisons by citizens, by testing for these reference point effects in managers.

In terms of substantive results, we find that social reference points matter. More specifically, the results suggest that comparisons with a peer significantly affect how managers interpret performance information, and consistent with our expectations, negative deviations from the social reference point matter more. Indeed, managers who were told that the performance metric they saw was in the bottom half of peer schools rated organizational performance significantly lower than the control group across both experiments. We did not find a significant effect for historical comparisons, regardless of whether performance was increasing or decreasing and regardless of whether historical shifts were paired with or administered as separate treatments from social comparisons. This result is a deviation from previous work on the interpretation of performance information by citizens and warrants further discussion, which we believe can help to illuminate the contributions of this study.

The first of these is the fact that this study investigates reference effects on the interpretation of performance information in a sample of *professional managers*. As noted above, this group is among the primary targets for such information and there is evidence that elites use information differently than citizens. As such, we might expect differences between the two groups and that is exactly what we find. For example, while Olsen (2017) and Charbonneau and Van Ryzin (2015) concluded that social comparisons probably play a stronger role in citizen evaluations than do historical comparisons, both studies still found some influence for the latter. We do not replicate that result in our pool of professional managers, suggesting that this group may place less emphasis on historical comparisons than do citizens.

119

This result has some potentially significant implications. First it suggests that studies of citizens may offer an incomplete picture of performance information use by professional managers and implies that the latter should be the subject of more research, despite the difficulties. Second, the results suggest that we may want to frame performance information in different ways, choosing different reference points, depending on the target group for that information.

The strong effect of social comparisons in our study draws attention to what we believe is another contribution. It is important to remember that previous comparisons of social versus historical reference points were drawn from separate experiments (Olsen 2017) or from experiments that could not accommodate all the potential points comparisons (Charbonneau and Van Ryzin 2015). Alternatively, the design in Experiment I allows us to directly compare the influence of social and historical frames. The confirmation of the relative importance of social comparisons in that design represents another contribution to this literature.

Finally, we believe that including managers from both the public- and private-sectors in our subject pool represents a contribution to both work on performance information use and to the longstanding debate on differences between these sectors. Interestingly, our results do not suggest consistent differences between public and private managers' uses of historical versus social reference points when interpreting performance information. In all but one case, the responses by managers from different sectors to different types of comparisons and different directional changes in performance were statistically indistinguishable. The only significant difference we observed was in the second experiment, where private sector managers were more responsive to *positively* framed information when making social comparisons than were their public sector counterparts. The relative *lack* of distinction between public and private managers

120

in processing performance information may mean that performance measurement and management systems may be more portable across sectors than previously thought.

Before concluding, we need to acknowledge some limitations of this study that suggest the need for replication and point the way forward for future research. We acknowledge that we need to be cautious regarding inferences about the power of historical comparisons because of the relatively small annual performance change (2% historical change) that we use as a treatment. As noted above, we chose this figure because it is a good approximation of the *average* performance shift that school's experience from year to year, but it may fall within the margin of indifference and as a result not be enough to move the manager to update any belief they may have had about the school's performance (Meier et al., 2015). Future experiments will manipulate the annual change parameter to better understand the size of this margin for historical comparisons and if the margin differs by policy area. Second, as noted above we did not ask our respondents to assess their own performance or the performance of the organization for which they work. Future work will attempt to better tie the information provided to subjects to their own work experience and will ask for personal, as well as organizational, assessments. Finally, respondents were asked to evaluate static performance information. While this was a limitation of sample size in this study, future work will randomize the specific performance information that subjects see (see Olsen 2017).

121

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review, 84*(6), 485.

Andersen, S. C., & Hjortskov, M. (2015). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory, 26*(4), 647-662. doi:doi: 10.1093/jopart/muv036

Andersen, S. C., & Moynihan, D. P. (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. doi:doi:10.1093/jopart/muv038

Anderson, D. M., & Edwards, B. C. (2015). Unfulfilled Promise: Laboratory experiments in public management research. *Public Management Review, 17*(10), 1518-1542.

Bækgaard M., Baethge C., Blom-Hansen J., Dunlop C. A., Esteve M., Jakobsen M., Kisida B., Marvel J., Moseley A., Serritzlew S., Stewart P., Thomsen, M. K., and Wolf, P. J. (2015). Conducting Experiments in Public Management Research: A Practical Guide. International Public Management Journal, 18(2): 323-342.

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

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

Baumeister, R. F., Bratslavsky, E., Finkenauer, C. and Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, 5(4): 323-370.

Behn, R. D. (1995). The big questions of public management. *Public Administration Review*, 313-324.

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

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

Blom-Hansen, J., Morton, R., Serritzlew, S. (2015). Experiments in Public Management Research. *International Public Management Journal*, 18(2): 151-170.

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

Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management, 29*(2), 110-131.

Boyne, G. A., James, O., John, P., and Petrovsky, N. (2009). Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments. *Journal of Politics*, 71(4): 1273-1284.

Charbonneau, É. and Bellavance, F. (2014). Blame Avoidance in Public Reporting: Evidence From a Provincially Mandated Municipal Performance Measurement Regime. *Public Performance & Management Review*, 35(3): 399-421.

Charbonneau, É., & Van Ryzin, G. G. (2015). Benchmarks and citizen judgments of local government performance: Findings from a survey experiment. *Public Management Review, 17*(2), 288-304.

Druckman, James N. 2004. Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects. *American Political Science Review,* 98(4): 671–86.

Festinger, L. (1954). A theory of social comparison processes. *Human relations, 7*(2), 117-140.

Freeman, R. A. (1954). How can voters understand? *National Civic Review, 43*(3), 123-129.

George, B., Desmidt, S., Nielsen, P., & Bækgaard, M. (2016). *Rational planning and politicians' attitudes to spending and reform: replication and extension of a survey experiment.* Paper presented at the 2016 IRSPM Annual Conference.

Hamilton, L. C. (2008). *Statistics with STATA: Version 10*. Belmont, CA: Brooks/Cole.

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.

Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review,* 9: 191–210.

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

Jacobsen, R. and Saultz, A. (2016). Will Performance Management Restore Citizens' Faith in Public Education? *Public Performance & Management Review*, 39(2): 476-497.

Jacobsen, R., Saultz, A., and Snyder, J. W. (2013). When Accountability Strategies Collide: Do Policy Changes That Raise Accountability Standards Also Erode Public Satisfaction? *Educational Policy*, 27(2): 360-389.

123

Jacobsen, R., Snyder, J. W., and Saultz, A. (2014). Informing or Shaping Public Opinion? The Influence of School Accountability Data Format on Public Perceptions of School Quality. *American Journal of Education*, 121(1): 1-27.

James, O. (2011a). Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government. *Public Administration*, 89(4): 1419-1435.

James, O. (2011b). Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments. *Journal of Public Administration Research and Theory*, 21(3): 399-418.

James, O., Jilke, S. R., and Van Ryzin, G. G., (2017). Causal Inference and the Design and Analysis of Experiments. In O. James, S. R. Jilke, & G. G. V. Ryzin (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 59-88). New York: Cambridge University Press.

James, O., Jilke, S. R., & Ryzin, G. G. V. (Eds.). (2017). *Experiments in Public Management Research: Challenges and Contributions*. New York: Cambridge University Press.

James, O. and John, P. (2006). Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments. *Journal of Public Administration Resarch and Theory*, 17: 567-580.

James, O. and Moseley, A. (2014). Does Performance Information About Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour? Field Experiments with Absolute and Relative Performance Information. *Public Administration*, 92(2): 493-511.

Jerit, J., Barabas, J., and Bolsen, T. (2006). Citizens, knowledge, and the information environment. *American Journal of Political Science*, 50(2): 266-282.

Jilke, S., Van de Walle, S., & Kim, S. (2016). Generating usable knowledge through an experimental approach to public administration. *Public Administration Review, 76*(1), 69-72.

Julnes, P. d. L., & Holzer, M. (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.

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice *Nobel prize lecture* (Vol. 8, pp. 449-489).

Kroll, A. (2014). Why Performance Information Use Varies Among Public Managers: Testing Manager-Related Explanations. *International Public Management Journal*, 17(2): 174-201.

124

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

Kroll, A. (2015b). Explaining the Use of Performance Information by Public Managers: A planned-behavior approach. *The American Review of Public Administration*, 201-215.

Leckie, G. and Goldstein, H. (2009). The Limitations of Using School League Tables to Inform School Choice. *Journal of the Royal Statistical Society: Series A*, 172(4), 835-851.

Levin, Irwin P., Sandra L. Schneider, and Gary J. Gaeth. 1998. All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes,* 76(2): 149–188.

Luskin, R. C. (1987). Measuring political sophistication. *American Journal of Political Science*, 31(4): 856-899.

Margetts, H. Z. (2011). Experiments for Public Management Research. *Public Management Review*, 13(2): 189-208.

Meier, K. J., Favero, N., and Zhu, L. (2015). Performance Gaps and Managerial Decisions: A Bayesian Decision Theory of Managerial Activity. *Journal of Public Administration Research and Theory*, 25(4), 1221-1246.

Mintz, A., Redd, S. B., and Vedlitz, A. (2006). Can We Generalize from Student Experiments to the Real World in Political Science, Military Affairs, and International Relations? *Journal of Conflict Resolution*, 50(5): 757-776.

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. In R. F. Durant (Ed.), *The Oxford Handbook of American Bureaucracy*. 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., & Hawes, D. P. (2012). Responsiveness to reform values: The influence of the environment on performance information use. *Public Administration Review, 72*(s1), S95-S105.

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

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

Moynihan, D. P., & Lavertu, S. (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., Nielsen, P. A., & Kroll, A. (2017). Managerial Use of Performance Data by Bureaucrats and Politicians. In O. James, S. R. Jilke, & G. G. V. Ryzin (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 244-269). New York: Cambridge University Press.

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

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

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

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*(3), 472-489.

Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General, 138*(1), 1-21.

Mussweiler, T., & Posten, A.-C. (2012). Relatively certain! Comparative thinking reduces uncertainty. *Cognition, 122*(2), 236-240.

Nielsen, P. A. (2013). *Performance Information in Politics and Public Management: Impacts on Decision Making and Performance.* (PhD), Aarhus University, Aarhus, Denmark.

Nielsen, P. A. (2013). Performance Management, Managerial Authority, and Public Service Performance. *Journal of Public Administration Research and Theory, 24*(2), 431-458. doi:doi: 10.1093/jopart/mut025

Nielsen, P. A., & Bækgaard, M. (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. doi:10.1093/jopart/mut051

Nielsen, P. A. and Moynihan, D. P. (2017). How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy. *Journal of Public Administration Research and Theory*, 27(2): 269-283.

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. (2017). Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations. *Journal of Public Administration Research and Theory*. doi:

Oppenheimer, D. M. and Kelso, E. (2015). Information Processing as a Paradigm for Decision Making. *Annual Review of Psychology*, 66: 277-294.

Rabin, Matthew. 1998. Psychology and Economics. *Journal of Economic Literature,* 36(1): 11–46.

Rozin, P. and Royzman, E. G. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, 5(4): 296-320.

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. A. (1939). The administrator in search of statistics. *Public Management, 21*, 106-109.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics, 69*(1), 99-118.

Skowronski, J. J. and Carlston, D. E. (1989). Negativity and Extremity Biases in Impression Formation: A Review of Explanations. *Psychological Bulletin*, 105(1): 131-142.

Stritch, J., Pedersen, M. J., and Taggart, G. (2017). The Opportunities and Limitations of Using Mechanical Turk (MTURK) in Public Administration and Management Scholarship. *International Public Management Journal*, 20(3): 489-511.

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

Tsang, E. W. K. and Kwan, K.-M. 1999. Replication and theory development in organizational science: a critical realist perspective, *Academy of Management Review* 24(4): 759-780.

Van de Walle, S., & Van Dooren, W. (2011). Introduction: Using Public Sector Performance Information. In W. Van Dooren & S. Van de Walle (Eds.), *Performance Information in the Public Sector: How it is Used* (pp. 1-12). New York: Palgrave Macmillan.

Walker, R. M., James, O., and Brewer, G. A. 2017a. Replication, experiments and knowledge in public management research, *Public Management Review* 19(9): 1221-1234.

Walker, R. M., Lee, M. J., and James, O. 2017b. Replication of Experimental Research: Implications for the Study of Public Management. In O. James, S. R. Jilke, & G. G. V. Ryzin (Eds.), *Experiments in Public Management Research: Challenges and Contributions* (pp. 439-460). New York: Cambridge University Press.

Wilson, D. and Piebalga, A. (2008). Performance Measures, Ranking and Parental Choice: An Analysis of the English School League Tables. *International Public Management Journal*, 11(3): 344-366.

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

Yockey, M. D., & Kruml, S. M. (2009). Everything is relative, but relative to what? Defining and identifying reference points. *Journal of Business and Management, 15*(1), 95-109.

Figure A1: Experiment I Vignette Workflow

Introduction (all respondents)	You will now be asked to respond to a question pertaining to educational performance data. For this question you are to assess the performance of an unnamed high school. For this question you will see the pass rate on a standardized **ENGLISH** exam. Administrators at the high school recently found out the results of yearly standardized testing. They want you to provide a performance assessment based upon the information you will see. Only use the information in front of you to assess each school's performance.				
Vignettes (random assignment)	_Control_	_Historical (Increase), Social (Top half)_	_Historical (Increase), Social (Bottom half)_	_Historical (Decrease), Social (Top half)_	_Historical (Decrease), Social (Bottom half)_
Vignette text	English Exam: **77**% of students in "High School A" passed their standardized English exam.	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10).	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% increase from the previous year. It also means the school was in the bottom half of local schools in the area (7th out of 10).	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% decrease from the previous year. It also means the school was in the top half of local schools in the area (3rd out of 10).	English Exam: **77**% of students in "High School A" passed their standardized English exam. This represents a 2% decrease from the previous year. It also means the school was in the bottom half of local schools in the area (7th out of 10).
Assessment prompt (all respondents)	Assuming this is the only information available to you, use the sliding scale (0-100) to <u>assess the overall performance of HIGH SCHOOL A over the last year:</u>				

Figure A2: Experiment II Vignette Workflow

Introduction (all respondents)	You will now be asked to respond to a question pertaining to educational performance data. For this question you are to assess the performance of an unnamed high school. For this question you will see the pass rate on a standardized **MATH** exam. Administrators at the high school recently found out the results of yearly standardized testing. They want you to provide a performance assessment based upon the information you will see. Only use the information in front of you to assess each school's performance.				
Vignettes (random assignment)	_Control_	_Historical (Increase)_	_Historical (Decrease)_	_Social (Top half)_	_Social (Bottom half)_
Vignette text	Math Exam: **79**% of students in "High School B" passed their standardized Math exam.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This represents a 2% increase from the previous year.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This represents a 2% decrease from the previous year.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This means the school was in the top half of local schools in the area.	Math Exam: **79**% of students in "High School B" passed their standardized Math exam. This means the school was in the bottom half of local schools in the area.
Assessment prompt (all respondents)	Assuming this is the only information available to you, use the sliding scale (0-100) to assess the overall performance of HIGH SCHOOL B over the last year:				

Table A1: Impact of Comparison Frames on Performance Assessment with Pre-Treatment Covariate

Regressions with Control Variables		
	English Exam (Experiment 1)	Math Exam (Experiment 2)
Treatment 1	0.23 (3.08)	-1.34 (2.45)
Treatment 2	-12.89*** (2.99)	-2.42 (2.48)
Treatment 3	-0.39 (3.00)	1.62 (2.50)
Treatment 4	-11.61*** (3.05)	-12.08*** (2.48)
Age	0.09 (.18)	-0.01 (0.15)
Education	-0.09 (.18)	0.68 (.40)
Female	0.52 (2.01)	0.60 (1.66)
Public Sector	-1.10 (1.97)	-0.36 (1.62)
Years in Workforce	-0.24 (.18)	-0.04 (.15)
Constant	77.36*** (5.37)	74.04*** (4.43)
N	297	297
R-squared	0.1346	0.1255
The outputs of this table provide the coefficients and standard errors from a regression of control variables in addition to the IV of interest (treatment group). The DV is the performance assessment indviduals responded for each experiment. *** significant at the 0.01 level.		

Exp I – Treatment 1: H(increase), S(upper half); Treatment 2: H(increase), S(lower half); Treatment 3: H(decrease), S(upper half); Treatment 4: H(decrease), S(lower half)
Exp II - Treatment 1: H(increase); Treatment 2: H(decrease): Treatment 3: S(upper half); Treatment 4: S(lower half)

130

CV
Sean Webeck
www.seanwebeck.com

Indiana University	Bloomington, IN
School of Public and Environmental Affairs	Phone: (512) 917-5625
Department of Political Science	Email: spwebeck@indiana.edu

EMPLOYMENT

Assistant Professor, Defense Resources Management Institute, Naval Postgraduate School, Monterey, CA (Accepted, position to begin Fall 2018).

EDUCATION

Ph.D., Public Policy, School of Public and Environmental Affairs and Department of Political Science, Indiana University. July 2018.

> *Committee:* Sean Nicholson-Crotty (Chair), William Bianco, Jill Nicholson-Crotty, Tom Rabovsky, and Asmus Olsen (Outside member)

> *Fields*: American Politics, Public Management, and Public Policy

> *Dissertation*: Applying a Behavioral Approach to the Use of Performance Information by Public Managers

MA, Security Studies, Kansas State University, August 2012.

MPA (focus in National Security Affairs), Troy University, October 2010.

BA, Political Science and Spanish, Wabash College, May 2003.

AA, Arabic, Defense Language Institute, November 2006.

PEER-REVIEWED PUBLICATIONS

Murdie, Amanda and Sean Webeck. (2015). "Responding to the Call: Human Security INGOs and Countries with a History of Civil War." *International Political Science Review*. 1-17. doi:10.1177/0192512114547372

Webeck, Sean and Sean Nicholson-Crotty. "The Use of Historical and Social Comparisons in Bureaucratic Decision Making." *International Public Management Review* (Conditionally accepted for publication)

SUBMITTED FOR PEER REVIEW

Nicholson-Crotty, Jill, Sean Nicholson-Crotty, and Sean Webeck. "Agency Mission, Pass-thru Grants, and the Implementation of Federal Programs in the American States." *Revise and Resubmit*

Nicholson-Crotty, Jill and Sean Webeck. "Nonprofit Advocacy Intermediaries: Competing with state actors to distribute federal grants." *Revise and Resubmit*

Webeck, Sean, Jill Nicholson-Crotty and Sean Nicholson-Crotty. "Negativity Bias, Gender, and Performance Information Use." *Revise and Resubmit*

BOOK CHAPTERS
Nicholson-Crotty, Sean and Sean Webeck. "What we Know about the Interaction between Politics and Administration in the Policy Making Process." *Handbook of Public Administration*, forthcoming.

BOOK REVIEWS
Webeck, Sean. "A Good Place to Start." A review of Christopher Pollitt's *Advanced Introduction to Public Management and Administration*. *Journal of Public Administration Research and Theory*, forthcoming.

WORKS IN PROGRESS
Nadella, Venkata and Sean Webeck. "Should I Stay or Should I Go: Veteran status and the determinants of turnover intention."

Nicholson-Crotty, Sean, Jill Nicholson-Crotty, and Sean Webeck. "Risky Choice in the Public Sector: An experimental test of framing effects."

Webeck, Sean and Venkata Nadella. "Exploring the Second-Dimension of Diversity: Managing Veterans in the U.S. Federal Government."

Webeck, Sean and Venkata Nadella. "Understanding Job Satisfaction and its Facets in U.S. Federal Government Employees: The case of Veterans."

Webeck, Sean. "The Role of Negative Information in Bureaucratic Decision Making: Understanding the interplay of framing, bias, and benchmarks."

CONFERENCE AND RESEARCH PRESENTATIONS
"The Use of Historical and Social Comparisons in Bureaucratic Decision Making." with Sean Nicholson-Crotty. Presented at the 2017 American Political Science Association Annual Conference (San Francisco).

"The Role of Negative Information in Bureaucratic Decision Making: Understanding the interplay of framing, bias, and benchmarks." Presented at the 2017 Public Management Research Conference (Washington, D.C.).

"Managing Veterans in the U.S. Federal Government: A diversity management perspective." with Venkata Nadella. Presented at the 2017 Public Management Research Conference (Washington, D.C.).

"Nonprofit Advocacy and Funding Intermediaries: Competing with state actors to distribute federal grants." with Jill Nicholson-Crotty. Presented at a symposium on nonprofit studies research at George Washington University (June 2017).

"The Role of Institutions in Bureaucratic Decision Making." An experimental research design presented at a panel entitled "Political Institutions and Elite Behavior 4: Experimental Approaches." 2017 Midwest Political Science Association's Annual Conference (Chicago).

"Negativity Bias, Gender, and Performance Information Use." Presented at a Masterclass on "Behavioral Public Administration" at the Royal Netherlands Academy of Arts and Sciences (*Koninklijke Nederlandse Akademie Van Wetenschappen*), Amsterdam, The Netherlands, 3-5 November 2016.

"Managing Veterans in the U.S. Federal Government: A research agenda." with Venkata Nadella. 2016 Midwest Political Science Association's Annual Conference (Chicago).

"Should I Stay or Should I Go: Determinants of Turnover Intention and Veteran Status." with Venkata Nadella. 2015 Public Management Research Conference (Minneapolis); Junior scholar pre-conference workshop, 2015 Annual Meeting of the American Political Science Association (San Francisco).

"Facets of Job Satisfaction in the U.S. Federal Government: The Case of Veterans." with Venkata Nadella. 2015 Midwest Political Science Association's Annual Conference (Chicago).

RESEARCH INTERESTS
Behavioral Public Administration
Bureaucratic Decision Making
Performance Management
Public Management
Veterans in the (Public-Sector) Workplace

GRADUATE TEACHING EXPERIENCE (INSTRUCTOR OF RECORD)

V263, Public Management – Indiana University – Fall 2015, Spring 2016, Fall 2016, Fall 2017

V336, Management Concepts and Applications II: Public & Private Organizations – Spring 2017, Spring 2018

TEACHING INTERESTS

American Politics
> Introduction to American Politics
> Institutions of American Politics (e.g., Bureaucracy, Congress, Presidency, etc.)
> Public Administration
> The Policy Process

Public Management
> Bureaucratic Decision Making
> Organizational Behavior
> Performance Management
> Theories of the Policy Process: National Security Policy

ADDITIONAL TRAINING

EITM 2017 (Empirical Implications of Theoretical Models), Berkeley, CA, Summer 2017.

FUNDING AND AWARDS

Institute for Humane Studies, Humane Studies Fellowship ($5000) – 2017-2018

Institute for Humane Studies, PhD Scholarship ($1500) – 2015-2016

Dr. Dana Mead Fellowship for interest in veterans' issues ($10,000) – 2014-15

Foreign Language and Area Studies (FLAS) (Turkish)
> Academic Year: 2014-15: declined
> Summer: 2013, 2014: accepted

Adam Smith Fellowship ($5,000) – 2013-2014

Boren Fellowship – Advanced Arabic Language Study in Amman, Jordan ($24,000) – 2011-12

Sheikh Ahmed Zaki Yamani Scholarship ($2,000) – 2011-2012

American Soldier Project Grant ($3,100) – 2010

GRANTS

Grant writer: Nicholson-Crotty, Sean (PI) and Sean Webeck (Co-PI). "Doctoral Dissertation
> Research in Decision, Risk, and Management Sciences: Applying a Behavioral Approach

to the Use of Performance Information by Bureaucrats." $15,210. National Science Foundation: Decision, Risk, and Management Sciences program. *Not funded*.

RESEARCH EXPERIENCE
Research Assistant, Drs. Sean and Jill Nicholson-Crotty, August 2014-August 2015

Research Assistant, Dr. Sameeksha Desai, SPEA, June 2012-July 2014

CAREER EXPERIENCE
Arabic Cryptologic Linguist/Analyst in the U.S. Army, April 2005-April 2010.
    15-month deployment to Mosul, Iraq (OIF 07-09).

Oral English Teacher, Hainan, China, September 2003-July 2004

LANGUAGES
Arabic (MSA – Intermediate-High); (Iraqi Dialect – Intermediate)

Mandarin Chinese (Novice)

Spanish (Intermediate-High/Advanced)

Turkish (Intermediate)

INTERNATIONAL EXPERIENCE (Professional)
Rwanda, May 2014, Teaching Assistant to Dr. Sameeksha Desai's class: Peace, Conflict, and Development

Amman, Jordan, December 2011-April 2012 (Boren Fellowship)

Mosul, Iraq, November 2007-January 2009 (U.S. Army deployment)

Haikou, People's Republic of China, August 2003-July 2004 (Oral English teacher)

Seville, Spain (and travels throughout Europe), July 2001-July 2002 (Junior year abroad)

Brussels and Lisbon, March 2001, Wabash College spring break immersion learning trip for a class on the Economics and Politics of the European Union