A THEORY AND EMPIRICAL EXAMINATION OF BUREAUCRATIC ERRORS

by

JUSTIN BENNETT BULLOCK

(Under the Direction of Laurence J. O'Toole)

ABSTRACT

For my dissertation, I have chosen to create and explore a theory of bureaucratic error. My theory of bureaucratic error seeks to determine the factors, at multiple levels that contribute to the occurrence of errors within government programs. The theory explores how factors at the program design, locus of implementation, office management, and bureaucratic decision-making levels interact within and across government programs to contribute to different types of government errors. This theory is tested empirically using federal program-level data, to examine across program variation in payment errors, and audit-level data, to examine variations in payment errors within the US Unemployment Insurance program. These studies find evidence suggesting that that the four levels identified in the theory of bureaucratic error contribute to the error generating process of payment errors within government programs.

INDEX WORDS: Public Administration, Theory, Bureaucrats, Decision making, Errors, Federal Programs, Unemployment Insurance

A THEORY AND EMPIRICAL EXAMINATION OF BUREAUCRATIC ERRORS

by

JUSTIN BENNETT BULLOCK

PhD, University of Georgia, 2014

MPA, University of Georgia, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PUBLIC ADMINISTRATION

ATHENS, GEORGIA

2014

A THEORY AND EMPIRICAL EXAMINATION OF BUREAUCRATIC ERRORS

by

Justin Bennett Bullock

Major Professor:	Laurence J. O'Toole
Committee:	W. Dave Bradford Hal G. Rainey Andrew B. Whitford

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2014

DEDICATION

For my parents, Mike and Valerie Bullock, who have always encouraged me to chase my dreams, and who always sacrificed to ensure that I had the best opportunity for a better life. And for my late grandparents, Bennett Morris and Verna Mae Morris, whom I always strive to make proud.

ACKNOWLEDGEMENTS

It is not possible to thank all of the individuals who have dedicated their time and energy in helping me throughout this process. However, I will mention a few individuals who have played a particularly large role in my life leading up to now. I would like to thank my parents for always supporting me in all my endeavors. I would also like to thank my dissertation committee – David Bradford, Hal Rainey, and Andy Whitford – for their time and input, without them this document would never have been completed. Special thanks to my major advisor and committee chair, Larry O'Toole, who guided me every step of the way and continually encouraged me to pursue the questions and ideas in this document. It is through his dedication and that of the committee, more broadly, that I was able to complete this dissertation. Thanks to Melody Herrington's tireless edits. She helped me to find my voice, time after time. Thanks to my department head, Ed Kellough, for protecting my time and making sure I had all the resources that I needed. Thanks to Jeffrey Wenger for serving as my first advisor and providing me with guidance to begin learning the craft of academic writing. Thanks also to Vicky Wilkins for her continual guidance, support, and for giving me my first opportunity (along with Jeffrey Wenger) as a co-author. Thanks to Langford Holbrook and Danny Bivins for always doing their best to promote my career, and for serving as two of my first professional role models. Thanks to the UGA public administration faculty for their continual support, encouragement, and guidance. Thanks to my fellow students in the UGA public administration department for all of the engaging and thought-provoking conversations. Also, thanks in particular to Justin Stritch who has always been there to encourage and support me every step of the way. Also thanks to all the conference

attendees who provided valuable feedback to earlier drafts of the chapters of this dissertation. Finally, thanks to my two best friends, Robert Clark and Matthew Hickman for their continual encouragement and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

A THEORY OF BUREAUCRATIC ERROR

Introduction

The goal for this essay is to develop a theory and corresponding typology for analyzing, evaluating, and testing determinants of different types of bureaucratic errors. Previous public administration literature has examined these issues using a piecemeal approach to focus on performance (Hvidman & Andersen, 2014), decision making (Avellaneda, 2012; Nutt, 2006), and errors (Fernandez & Moldogaziev, 2013; Lens, 2009; Zajac, 1996). Other disciplines, however, have developed entire sub-specialties for understanding decision-making processes (Bazerman & Moore, 2008; Bellman & Zadeh, 1970; Cyert & March, 1963; Janis & Mann, 1977; Klein, Orasanu, Calderwood, & Zsambok, 1993; Plous, 1993; Zeleny & Cochrane, 1982) and tracing how they are affected by such factors as organizational context (Busenitz & Barney, 1997; Jones, 1991; Simon, 1979), individual characteristics (Trevino, 1986), and incentive structures (Groves, 1973; Laffont & Tirole, 1991; G. J. Miller & Whitford, 2007; O'Reilly, 1983; Tosi, Katz, & Gomez-Mejia, 1997). This theory will build upon these and other literatures, producing a holistic discussion of errors found in government programs.

This work addresses how factors such as program mission, organizational design, varying incentive structures, highlighted performance measures, individual characteristics of clients, and attitudes and values of bureaucrats might affect the propensity of bureaucrats to make various types of errors, ranging from purposive errors to unintentional errors. This paper takes a broad approach in considering how these various factors intermingle to contribute to a wide variation

of the presence of errors in government programs. Often these variables are analyzed and discussed in isolation from one another, but I contend that these variables should be taken in concert to address the presence of bureaucratic errors.

In this context, what is meant by the term "error"? This study will define and categorize specific types of errors that may occur within a public organization, illustrate them with real-world examples, and survey the data available to empirically examine them. Once the various types of errors are defined, I will discuss the different levels at which an error might occur (program design, locus of implementation, office management, and administering by street-level bureaucrats). From here, I will explain why a particular type of error (payment error) may be useful for studying and evaluating this theory. Then, I discuss two examples of government programs (Medicare Fee-for-Service and Unemployment Insurance) that might benefit from applying the theory of bureaucratic errors in a critical way. And, finally, the essay will conclude with a discussion of the importance of this area of inquiry, and the care which should be taken when examining these types of questions.

What is an Error?

So, what is an error? *An error is any deviation from an intended outcome that is mandated by either law or organizational rules*. This definition is purposefully broad. An error can be almost anything. An error can be a clerical error of a field agent within the US Census Bureau while entering data on a household from Des Moines, Iowa. An error can be the recording of an incorrect diagnosis code made by a Medicare Administrative Contractor who has contracted with Medicare to efficiently and effectively deliver Medicare processing to the constituents in some region of the country. An error can be made by an Unemployment Insurance agent in determining the eligibility of a claimant for unemployment insurance benefit.

An error can be made in the application process of the Food and Drug Administration when examining the safety of a pharmaceutical. An error can be made by the Social Security Administration in their calculation of benefit amounts as a result of bad income information. An error can be made by the Internal Revenue Service in their determination of a taxpayer's tax liability. Not only can an error be almost anything and occur in any public program, but the errors may also be large or small. They may be systematic or they may be random. They may be persistent or a one-time occurrence.

An understanding of the nature of different types of errors can determine its level of detriment to program performance, and the degree to which it negatively affects the overall efficiency or effectiveness of government as a functioning unit. Thus categorizations of these errors are useful from a theoretical standpoint. One valuable categorization of errors is the *intent* behind the error: whether or not the error is intentional (which for illustrative purposes I will call "purposive errors") or unintentional (which I will call "accidental errors"). A second useful categorization is *outcome*, or whether or not the error is favorable or punitive to some party. These categories will help us examine how these errors will affect a variety of parties.

Error Categorization

Purposive vs. Accidental

As mentioned above, one useful categorization of errors is whether or not the error was committed in a purposeful way (purposive) or in an accidental manner (accidental). Table 1.1 displays this categorization.

What is a purposive error? A purposive error is an error (recall an error is any deviation from an intended outcome that is mandated by either law or organizational rules) which is made in a *deliberate* and *intentional* manner. A purposive error is made consciously and intentionally

by a bureaucrat to serve some purpose. These purposive errors are made to either favor or punish (the second categorization) a variety of parties, such as clients, other bureaucrats, the program mission, or the political party in power. This categorization (taken with the *outcome* of the error) gives us eight types of purposive errors. These include client favorable, client punitive, other bureaucrat favorable, other bureaucrat punitive, mission favorable, mission punitive, political-power-in-party favorable, or political-power-in-party punitive.[1] Thus a purposive error may result from favorable treatment of both a client and of the overall mission of the bureaucrat.

What about accidental errors? Accidental errors are errors (and, once again, recall that an error is any deviation from an intended outcome that is mandated by either law or organizational rules) which are made unintentionally and should not be systematic in nature unless as result of poor policy design, work flow operations, or subconscious cognitive biases. Accidental errors arise from processes that are *unintentional*, but benefit or punish the same sets of groups (clients, other bureaucrats, mission, or political party in power). While both purposive and accidental errors affect organizational outcomes and performance, they do so according to differing rules. One, accidental errors in aggregate should not systematically affect specific groups, unless the accidental errors are a result of a flawed work process, or of some subconscious process, such as a cognitive bias. Second, on average, one can assume that accidental errors are likely to occur less frequently and be smaller in magnitude (again, unless there is some work flow or policy design flaw).

Favorable vs. Punitive

A second dimension across which errors can be categorized is whether or not the error has favorable or punitive impact on some party. An error is favorable if the affected party

[1] It is important to note that these categories are neither mutually exclusive nor exhaustive but rather serve as examples of possible categorizations of bureaucratic errors.

receives some positive treatment, over and above what is mandated by either law or organizational rules. For example, a bureaucrat may award some client benefits for which he or she is not eligible, or more benefits than legally appropriate. Alternatively, this bureaucrat may take actions that further a program's mission, but do so outside of applicable legal or organizational rules.

On the other hand, an error is punitive if the affected party receives some negative treatment, or lack of treatment that is mandated either by law or organizational rule. For example, if a bureaucrat denies benefits to a client who is legally entitled to them, then the error is punitive. A punitive bureaucratic error might negatively affect a particular political power. The roll-out of the Affordable Care Act (ACA) is one apt illustration. Errors made in the development of the ACA website not only punished clients who were seeking the benefits, but also proved to be punitive to President Obama's political influence, as his favorability took a major hit (Dutton, De Pinto, Salvanto, & Backus, 2013).

Affected Parties

There are at minimum four parties affected by bureaucratic errors: 1) clients, 2) other bureaucrats, 3) program mission, and 4) political principals. Clients are most often directly affected by bureaucratic errors. If a bureaucrat makes an error that affects the receipt of benefits or the amounts of benefits, this directly affects the client. Other bureaucrats can also be affected. If the error made has group components, or involves co-workers up or down the hierarchical line, the error affects these individuals as well. The program's mission can also be positively or negatively affected by bureaucratic errors. Bureaucrats may engage in behavior outside of legal mandates and/or organizational rules, and this behavior may positively or negatively affect the

mission of the program.[2] Finally, political power held by political agents may also be affected as a result of bureaucrats' errors. These errors can either enhance a political agent's power or decrease it.

With these categorizations in mind, I now turn to a discussion of the ways in which these errors may be "built into" these programs through their design, implementation, management, and bureaucrat values.

Levels at Which Errors are "Built Into" Government Programs

For this theory, it is important not only to think about how errors are *categorized* and the *effects* they may have on certain groups, but also to consider how the many different facets of a given government program may affect the likelihood of errors occurring within that program. If constructed correctly, these various levels will help to identify why performance differs both *across* and *within* government programs. There are four levels, or "nests," in which errors might be "built into" government programs: 1) program design, 2) locus of implementation, 3) office management, and 4) administering by street-level bureaucrats. These strata not only organize the various factors influencing program performance, but they also represent a hierarchy, within which each of the trailing levels builds on the prior, and thus directly influences, the following level. The program design contains the framework for the locus of implementation, and the locus of implementation influences the type of management, level of office-level resources, and the incentives, limitations, and values of the street-level bureaucrat. Figure 1 illustrates the four levels and how they are interrelated.

[2] The bureaucrats purposefully engaging in these errors and are referred to as rogue agents (Wenger & Wilkins, 2009).

The following sections will discuss, in detail, how decisions made at each of these levels, along with how additional external factors at each level, affect the performance and the propensity for errors *within* and *across* government programs.

Program Design

It should not come as a surprise that the design of a government program would affect the likelihood of that program to have more or fewer errors (S. H Linder & Peters, 1984; Stephen H. Linder & Peters, 1987; Laurence J. O'Toole, 2000). This is the level at which both policy makers and legislators may need to take the most caution, and at which much variation across programs may be explained. The design of government programs affects every following level in an unforgiving way (Lynn, Heinrich, & Hill, 2000; May, 1992; Soss, 1999). There are five major factors to be considered with the design of a program: discretion, goals, information, unintended consequences, and funding mechanism.

The first of these five factors is discretion. By discretion, I mean the level of decision-making power over program implementation and administration enjoyed by lower-level policy actors (Handler, 1992; Keiser, 1999; Lipsky, 1980). In other words, how much "say" do bureaucrats have in who receives benefits from a government program, how much they receive, and the manner in which they receive it. Additionally, how much can bureaucrats influence the rulemaking of the program? Can they make serious alterations in how resources are distributed?

The more discretion given to bureaucrats over program implementation and administration, the more likely the program is to experience more errors (Christensen, Szmer, & Stritch, 2012; Wenger & Wilkins, 2009). Why is this? Is this an indictment of the competency of bureaucrats? No. This simply notes that increased discretion on behalf of the bureaucrats creates more *room* for errors to occur. It is important to note the tradeoff here. While more discretion

does allow more room for errors, it also allows programs to have more flexibility and to more easily address complex and unique situations that may arise within the confines of a program. Thus, as I will continue to iterate throughout this discussion, these decisions are almost always made by considering tradeoffs; in this case, the tradeoff is made between errors and flexibility.

The second of the five factors that affect the influence of program design on bureaucratic errors is goals. Goals of government can be thought of across two major lines, clarity and politicization. The United States government has engaged in attempts to identify whether the goals of a program are clearer or more ambiguous. There is a clear connection between goal clarity and the expectation of bureaucratic errors (Chun & Rainey, 2005a, 2005b; Jung, 2011). If the goals are relatively more ambiguous in nature, then there is a higher likelihood of increased numbers of bureaucratic errors. If it is not particularly clear to the bureaucrat what the goal of the program is, and thus how it should be implemented, then the propensity for errors is obviously higher.

The politicization of goals can also influence the likelihood of bureaucratic errors within a program. When an agency or program's goals are perceived by the public to be political in nature (for example, the Affordable Care Act or the Consumer Financial Protection Bureau), the bureaucrats are more likely to be influenced by external pressures affecting the propensity of errors (Fernandez, Cho, & Perry, 2010; Gilmour & Lewis, 2006; Lewis, 2008). On the contrary, if the goals are less political in nature, the bureaucrats are less likely to be influenced by external political factors, and thus less likely to commit errors. This could also be viewed under the lens of financing and uncertainty as well. Programs with politicized goals are more likely to see fluctuations in resource availability as political climates vary, yielding an increased likelihood of errors related to variability and uncertainty in resources.

The third of these five factors is information, and, particularly, information gathering. The proposition is not that information gathering, in and of itself, is directly related to bureaucratic errors. Rather, the amount and types of information that must be gathered by bureaucrats while under the constant constraint of limited time and resources may result in errors. If eligibility requirements or benefit formulas are complex and require the intensive gathering of information from multiple sources that range in their level of objectivity, then this is a high burden for information gathering, thus increasing the likelihood of bureaucratic error. Again, this is not meant to suggest incompetence of behalf of bureaucrats. Instead, I argue that the more demands on information, the more *room* for errors to occur.

A program in which the information collection process is straightforward and routinized is less likely to experience errors than a program that gathers information in complex and varied circumstances. Again, simple, routine circumstances yield a lower propensity for bureaucratic errors, while complicated, non-routine circumstances result in a higher propensity for bureaucratic errors (Wenger and Wilkins, 2009). However, some societal problems demand complex solutions, and thus require a balancing act between the propensity for errors and the need for appropriately complex solutions. This tradeoff repeats continually, but the influence of some of these factors may be muted or increased by decisions made at lower levels.

The fourth factor is unintended consequences. As economists have suggested for years, government intervention has a distorting effect on the marketplace and thus, often, alters the incentives of private actors (G. Miller, 2000; Pauly, 1974) . This is not the place to give a full treatment of unintended consequences as a result of government programs, but it is important to note that these programs can introduce both adverse selection and moral hazard into the private actors' decision-making calculus. The degree to which a program introduces these dangers into a

decision maker's calculus relates to program efficiency, and thus the occurrence of program errors. Finally, the program design may create perverse incentives for the *bureaucrat's* behavior and work effort, but this detailed discussion will be saved for a later time.[3]

The fifth and final program design factor to be considered is the funding mechanism selected for a program. The type of funding chosen can affect the propensity for program errors in several ways (Sandfort, Selden, & Sowa, 2008). One should keep in mind both the funding mechanism for the administration of a program and the funding stream of the actual payments, as these processes are sometimes separate. One path by which the funding mechanism can affect the propensity for errors is the level of predictability of the budget allocation. If the program is under constant strain of resources, and the budget for the next year is particularly unknown, this can lead to large amounts of uncertainty, and thus increase the likelihood of errors. Additionally, extended resource deprivation leads to an increased likelihood of error. Finally, if program bureaucrats are constantly concerned over the availability of resources, they may be forced to appeal to special interest groups and lobby mechanisms to help garner resources. Under this scenario, bureaucrats' conflict of interest increases the likelihood for errors in program delivery. Finally, the funding mechanism itself may create perverse incentives, or, if constructed well, create incentives for efficiency and effectiveness.

While these five factors are not exhaustive, they do highlight several important considerations for the relationship between program design and bureaucratic errors, and the tradeoffs associated with program design between factors such as bureaucratic errors, usefulness of discretion, goal clarity, politicization of goals, complexity of information gathering, and

[3] Government programs may also create perverse incentives for firms. These perverse incentives could encourage the private firms to attempt to "game" the system, thus making it more difficult for bureaucrats to administer a program. This would lead to a higher propensity for bureaucratic error within these programs that create perverse incentives for private actors.

distortionary impacts on the market and their associated unintended consequences. I now move the discussion to the ways in which these factors might influence the lower levels of locus of implementation, office management, and administration by street-level bureaucrats.

Locus of Implementation

The second level of potential structural error is the locus of implementation. I now narrow the focus to centralized vs. decentralized implementation. Programs, at whatever level they were conceived and designed, can be implemented at different levels of government. This is particularly true with American federalism. The federal government may pass a statute that I leaves to the states to implement. Additionally, states may pass laws that they leave to the local governments to implement. The main discussion will be the tradeoffs experienced by the governing body ultimately responsible for implementation. This framework can also apply if a governing body chooses to contract out the implementation of a government program to multiple private entities.

A centralized-versus-decentralized approach is characterized in four particular ways that affect the propensity for bureaucratic errors. First, in a centralized program, there is less variation among the implementing units (states, local governments, or private entities). If the federal government implements and runs a major program, the variance among the states can be expected to be less than if the lower units were implementing the program. This is neither a positive or negative statement, but rather an acknowledgement that decentralized governing units have more variation in the amount of errors than if the program were implemented by a single centralized governing unit.

Second, not only is there less variation if a program is centrally implemented, but central implementation may result in greater economies of scale, from one, larger governing body

pooling its resources to implement a program. The federal government, for example, is larger, and likely to have more in-house expertise, specialization, and internal resources than fifty independently operating states. This suggests that with lower economies of scale, decentralized implementation may lead to higher amounts of errors. Competing with this, however, is that the economies of scale might come at the cost of specializing implementation for specific contexts. The tradeoff is flexibility to adapt to the contextual environment.

Third, if a program is centrally implemented, then there will be less experimentation in the implementation of a program. This can be seen as either a positive or negative. One oft-cited reasoning for state implementation is that the states can provide a variety of opportunities to experiment with the "tinkering" of implementing a program (Beaton, 2008; Gardner, 1995; Greenberg, Linksz, & Mandell, 2003; Rabe, 2007). These states may then learn and adapt from both the best and worst practices of other states. Thus, to the degree to which states do differentially implement some programs, and furthermore, to the degree to which they learn from each other's mistakes and successes, experimentation should lead to wide variation in bureaucratic errors (particularly in the early years of a program's existence), but to lower bureaucratic errors over time, as the states implement best practices and decrease the overall likelihood of bureaucratic errors.

Fourth, and finally, the implementation method influences the degree to which political actors (who enacted the program) and policy designers (who designed the program) affect the bureaucratic errors within a program. The argument here is straightforward: the more removed the implementation of the program from the political actors and the policy designers, the less likely it is that these individuals can meddle with the program and thus interfere with its

implementation. This interference, by disrupting program administration, might increase the likelihood of bureaucratic errors.

These first two levels (program design and locus of implementation) build the basis on which I will now analyze the following two lower levels. The higher levels set the stage, and thus the parameters, within which the program will operate. (Thus we might call the first two levels the *parameters*.) Within these parameters, it is likely that a wide variation of bureaucratic errors have been "built into" the program. The lesser two levels will help to either *mitigate* or *aggravate* these structural bureaucratic errors. Thus, we might think of the following two levels as the *mitigating* levels.

Office Management

Public administration and public management scholars have spent a great deal of ink examining a variety of factors at the office management level (Rainey, 2009). This research often seeks to identify some archival or subjective variable as a measure of office performance, and then examines a host of variables that might predict the variance in office performance. Meier & O'Toole have contributed greatly to this literature in their examination of Texas school districts (Meier & O'Toole, 2002). Others have looked at the pass rates in schools in other countries (Hvidman & Andersen, 2014). Furthermore, scholars have also examined the factors that affect hospital performance (Johansen & Zhu, 2014). However performance is defined at the office (also known as organization) level, this research has focused on what management factors influence these outcomes.

Office management factors fit into two broad categories: management factors and resource factors. Management factors include incentive structures (including measuring

performance), work flow systems, networking activities, and leadership capabilities. Resource factors include labor supply, budget, and demand for services.

The first management factor is the design and implementation of incentive structures. This literature has a rich intellectual history (Clark & Wilson, 1961; Downs, 1964; Niskanen, 1971). Other authors, such as (Blau, 1969), have discussed the importance of incentive structures and phenomena such as "creaming" and "maladaptive behaviors." This research has seen a resurgence in recent years with work from Soss, Fording, & Schram (2011). It may not be possible to overstate the importance of incentive structures on office performance. There are rich generic management literatures following the effects of different incentive systems on workers. Additionally, many scholars have developed rich literatures on how these incentive structures and motivations differ from those in the private sector (Rainey & Bozeman, 2000; Rainey, 1983). The previous research emphasizes the importance of creating an incentive structure that correctly motivates bureaucrats to perform in a way that minimizes errors. There are no straightforward hypotheses explaining how incentive structures will affect the incidence of bureaucratic errors within an office. However, it seems quite obvious that a poorly constructed incentive system will lead to higher propensity for bureaucratic errors, while a properly constructed incentive system will lead to a lower propensity for bureaucratic errors.

The second management factor is workflow systems[4], or systems that have been put into place to organize the work that an office must accomplish. Work flow systems are studied closely in the business literature (Baron & Bielby, 1980, 1984; Frenkel, 1999), but less so in the public management literature. However, they remain very important. . What is the hierarchical structure of the office? How many subordinates report to each manager? How broad is the scope of authority for each manager? What are the checks in the system to make sure errors are caught?

[4] This is also known as the organization of work.

How are the departments of the office organized? These are just a few of the questions that surround work flow systems. It is also important to note that some offices that administer government programs have limited freedom to organize themselves, depending upon the legislation and the implementation directives.

A few authors have commented on the organization of work flow systems, if only indirectly, and under the guise of a different topic. (Bendor, 1985) discusses the importance of redundancy in government. He argues that we need redundancy in implementation of government programs in order to minimize the amounts of mistakes in the delivery of these services. Before Bendor, (Landau, 1969) also discussed the importance of redundancy in government. Following Bendor, (Chisholm, 1989) spoke to the problems of hierarchy, and discussed how coordination is possible without strict, multiple levels of hierarchy. These literatures highlight the importance of work flow systems *within* an office, but another important office variable discusses the importance of dealing with factors *outside* the office.

The third management factor is networking activities. Meier & O'Toole pioneered the examination of the importance of managerial networking in office performance (Meier & O'Toole, 2003, 2005; O'Toole & Meier, 2004). Their work illustrates the networked nature of implementing government programs: these offices do not work in silos. The managers who are effectively engaging in networking activities are gathering the information they need to efficiently implement their respective programs. The more effective the networking, the less likely the bureaucrats are to make errors, and the better the performance of the office.

The fourth management factor is leadership capabilities. While the three previous management factors discuss the importance of a variety of systems, structures, and behaviors, none of these specifically addresses the importance of leadership. The public administration

literature on leadership lags behind this topic in the generic literature (Wart, 2003). For a variety of reasons, the quality of leadership within an organization can directly impact the propensity for that organization to commit errors. The causal reasoning is twofold. First, better leaders are better able to "steer the ship." In other words, they can more effectively manage all of the processes involved in the production function of some government service or program. Second, better leaders are also better at motivating employees to do their jobs more effectively and efficiently, thus leading to fewer errors. *Properly* motivated workers are less likely to commit errors. For the likelihood of errors to decrease, workers must be motivated to specifically carry out the letter of the law. This is important, because as will be discussed later, bureaucrats may be motivated to "take matters into their own hands" and give certain groups of citizens more or less access to a program than that which they should receive according to the law. The merits, or demerits, of street-level bureaucrats using their discretion to limit or over-provide benefits to specific types of clients is a discussion for another place; however, one should note that if street-level bureaucrats are motivated to provide (and indeed do provide) these differential levels of treatment to different groups of individuals, these differential levels of treatment are, by definition, errors. Thus if bureaucrats are improperly motivated, they may be more likely to make purposive errors.

While management factors are important to the variation in office management, resource factors also play a critical role. These resource factors include labor supply, office budget, and demand for services. These factors are well-documented in other disciplines (such as economics and generic management) but often neglected in the discussion of public management and implementation.

The first of the three resource factors is labor supply. If a qualified supply of labor is not available to an office, then the performance of that office will suffer and the likelihood of errors will be greater. This can be a result of two phenomena. The office may be located in an area with a weak (relatively unskilled or small in numbers) labor supply, with the resulting lower wages insufficient to motivate qualified workers. Alternatively, the office may face relatively high competition for a given labor pool, and for this reason cannot induce qualified workers to work there. However, under either of these scenarios, the bureaucrats working in an office may not be qualified for the task for which they were hired. Thus the likelihood of errors within the office increases.

The second of the resource factors is the office budget. The office budget encompasses both the funding mechanism for the program and the method of resource allocation to the various offices. What is the method by which the various offices' (particularly when there are hundreds, if not thousands of offices) operating budgets are set? And does this formula create adverse incentives for the office, which lead to an increase in the likelihood of errors? For example, if the budget is a function of the number of claims an office files, or the number of clients served, these formulas may create perverse incentives on behalf of the office and bureaucrats.[5] These perverse incentives possibly increase the likelihood of errors occurring within an office.

The third of these resource factors is demand for the delivery of office services. In this context, demand is defined as the number of clients and citizens who are either entitled to or apply for the benefits or services of some particular program. This demand on behalf of clients and citizens can fluctuate as a function of a variety of variables: market, changing demographics of an area, external shocks, and others. If there are positive shocks to demand, or positive overall growth in demand that is not met, whether because it is not recognized by the budgetary formula

[5] An example is "creaming" (Blau, 1969).

or because the political atmosphere will not allow an increase in office resources, then an increase in demand is likely to be associated with an increase in the likelihood of bureaucratic errors. The logic is simple. If offices are continually expected to do more with the same amount of resources, then at some point the demand-to-resources ratio becomes too high, and the likelihood of errors within that office increases.

Administering by Street-Level Bureaucrats

The previous three levels all work to influence the behavior of street-level bureaucrats. The above factors set forth in both the parameter levels (program design and locus of implementation) and the mitigating level (office management) create the system in which street-level bureaucrats operate. Despite the major role the other levels play in influencing the behavior of street-level bureaucrats -- and thus their propensity to commit errors – they cannot fully explain them. Below I discuss literatures from three areas that can be utilized to examine factors at the bureaucrat level that may affect their likelihood to commit errors: decision-making limitations, cognitive biases, and bureaucratic values.

Research on decision-making limitations has a rich intellectual history in the literatures of psychology, generic management, public administration and public management. Simon highlighted two very important concepts for decision-making theorists: individuals are not perfectly rational decision makers, and, given that they are not perfect, they often *satisfice.* For Simon, satisficing suggested that decision makers have limited abilities to fully evaluate and properly weight all available options – and thus be fully rational -- for a given decision, and thus they optimize (or "maximize") decisions to the best of their ability (Simon, 1947). Following Simon, two psychologists, Daniel Kahneman and Amos Tversky, demonstrated specific ways in which individuals, under conditions of uncertainty, deviate from rationality (Daniel Kahneman &

Tversky, 1972; A. Tversky & Kahneman, 1974; Amos Tversky & Kahneman, 1971). The result of these experiments was Kahneman and Tversky's Prospect Theory (D. Kahneman & Tversky, 1979). Arguably the most important takeaway from this theory was the suggestion that under conditions of risk and uncertainty (conditions almost constantly found within organizations, both public and private), individuals are not always risk averse. However, when the decision is framed as a possible loss, individuals become risk-seeking in order to avoid a loss, and are thus considered to be *loss averse*.

Loss aversion, in particular, is an important concept for considering bureaucratic errors. It suggests a systematic tendency for bureaucrats to engage in behavior that prevents losses to them. These losses can be defined in a multitude of ways. Bureaucrats might avoid behavior that could negatively affect their reputation, their emotional state, or their wallet; they may make errors in an attempt to avoid these perceived losses. The literature on loss aversion, stemming from prospect theory, has helped to develop another line of literature that examines other types of systematic errors in decision making: cognitive biases.

The second area, cognitive biases, has been the focus of research for many years by Max Bazerman (Bazerman & Moore, 2008). Bazerman expanded on the ideas put forth by Khaneman and Tversky and identified several ways in which individuals make cognitively-biased decisions. Bazerman suggests that not only are decision makers limited in their time and resources, but also that these constraints result in shortcuts to decision making (as Simon suggested). These shortcuts may be biased in the form of loss aversion (as suggested by Khaneman and Tversky), but there may also be other systematic tendencies in these shortcuts (also known as heuristics). Some of these biases include overconfidence, ease of recall, insensitivity to sample size, regression to the mean, insufficient anchor adjustment, and confirmation trap (Daniel Kahneman,

2011). Studies suggest that these cognitive biases can greatly influence decision making and thus generate errors (Daniel Kahneman, 2011).

This literature has been less influential in public management. However, two recent studies have examined the presence of cognitive biases in public managers. (Moynihan & Lavertu, 2012) examined the propensity for election administration officials – who had a preference to technological advances – to adopt new voting technologies over and above their perceived value. This finding suggests a cognitive bias in favor of technology. These biases could create systematic errors in bureaucrats' decisions, as is shown by (Ryu, Wenger, & Wilkins, 2012). In this study, they postulate – with evidentiary support – that street-level bureaucrats working within the U.S. unemployment insurance program are influenced by cognitive biases in determining unemployment insurance eligibility.

Finally, not only is bureaucratic behavior affected by decision-making limitations and decision-making cognitive biases, but also by the values of the bureaucrats. This is a concept which has received a great deal of attention in the public administration literature. Many scholars have contributed to the representative bureaucracy literature, which examines how bureaucrats, both actively and passively, provide differential treatment to different types of clients (Riccucci & Meyers, 2004; Thompson, 1976; Wilkins & Keiser, 2006). Bureaucrats are more likely to provide special treatment to clients who resemble them. The construct undergirding this may be that bureaucrats either relate more easily to those who look like them, or they find those who resemble them to be more deserving of benefits (Mosher, 1968).

A number of scholars complement the representative bureaucracy literature by delving further into how street-level bureaucrats make decisions about who deserves benefits. One landmark study is a book by (Maynard-Moody & Musheno, 2003), which looks at how police,

teachers, and counselors decide how to allocate their time and resources, based upon who they view as deserving. Another strand of literature examines how beliefs about a just world might affect bureaucrats' beliefs about who deserves government benefits (Bullock, Wenger, & Wilkins, 2014). Further insight into deservingness can be found in (Schneider & Ingram, 1993). Their theory of the social constructions of target populations argues that those constructions affect the level and type of government policies directed at such populations. This can also be thought of as a way in which street-level bureaucrats think about providing assistance to different target populations requesting government benefits (Bagwell and Bullock, 2013).

Another separate, but rich, literature focuses on the motivators of bureaucratic behavior. Much work has been done on how bureaucrats are motivated, and how they are motivated differently from private-sector workers (Perry & Rainey, 1988; Rainey & Bozeman, 2000). This literature attempts to identify which types of motivations lead to things like increased performance and job satisfaction. The literature has, in general, found that a wide variety of motivations matter for performance, and that they differ as between public and private employees. Thus differing levels of motivations may affect the propensity for a bureaucrat to make errors.

Measuring Bureaucratic Errors: The Payment Error & The Auditor

This theory lays out a framework for thinking about the design, implementation, and management of government programs, with an eye to understand how one might minimize the errors found in these programs. While it is unlikely that any single theory can account for all of these factors, this theory offers a potentially useful approach that is as exhaustive as possible, predicts future program performance, and creates targeted approaches for improving the efficient delivery of government services. This theory, however, does create issues for researchers: what

kinds of errors should be evaluated? Who determines what constitutes an error? Here I will argue that while there are legitimate concerns about the role of auditors in this process, payment errors have unique strengths for studying bureaucratic errors.

Why use the payment error as the measure of bureaucratic error? While this type of error is limiting in some ways, it offers at least four distinct advantages for empirical researchers: 1) payment errors are clearly defined (thus error rates can be compared across multiple programs); 2) payment errors are a very important type of error with regards to the efficiency and effectiveness of government; 3) payment errors are usually made at the individual-bureaucrat level and thus can capture the end result of decisions made at all four levels; and 4) data are available.

The first advantage of focusing on payment errors is the simplicity with which they are defined. Often in government programs there are rules and regulations governing not only who receives payments, but also the amounts they should receive. Any deviation from these rules and regulations is considered to be a payment error. This clarity helps to mitigate concerns about consistently defining an error. This is not to say that we should not also be concerned with errors that may be made by auditors (as will be discussed below), but it does suggest that there should be relatively less concern for the role of auditors when the error is more clearly defined. Additionally, this clarity allows researchers to compare error rates across a wide variety of government programs on a similar metric. Previous measures in the literature rely on either subjective measures of performance or a measure that is specific to only one type of organization.

Secondly, payment errors are ideal for this study because they have serious ramifications for public administration. Payment errors, in particular, undermine the legitimacy and faith that

the general public has in the ability of government to perform effectively and efficiently. If the public can see large amounts of overpayments, then they may be frustrated by the waste of the program; if they see large amounts of underpayments, then they may feel cheated. These concerns strike to the very heart of administration, implementation, and government's ability to provide services.

The third advantage of using this phenomenon to study bureaucratic errors is that decisions about payments to clients are often made on the ground by street-level bureaucrats. This means that researchers can study variation across all four levels. Street-level bureaucrats are housed within a specific office, with a specific locus of implementation, and within a certain program. If payment decisions were made at higher levels, then researchers would not be able to test variation in the lower levels.[6] Fortunately, however, we have variation at all four levels, so we can create research designs that allow for studying variation either among all four levels (program design, locus of implementation, office management, and street-level bureaucrats) or, that allow us to focus in on one or two specific levels, to better understand the variation in those specific strata.

The fourth reason, and one that is particularly crucial for researchers who are empirically-minded, is that data are available to examine payment errors. By federal statute, going back to 2002,[7] the Office of Management and Budget (OMB) has required approximately 100 federal programs to systematically collect data on the amounts of improper payments, and the OMB makes this data publicly available. This provides researchers with a dataset with sufficient observations to conduct large-N multivariate analyses. Also, multiple federal programs make

[6] This feature allows for examining the variation both *across* and *within* programs.
[7] The formal name of the statute is "Improper Payments Information Act of 2002." Very similar pieces of legislation were also passed in 2010 and 2012 in an attempt to further focus and clarify the initial legislation.

data available through publicly published reports, and more data may be available through the Freedom of Information Act or other avenues.

While the payment error as a measure has significant strengths, one important concern is the role auditors play in collecting these data. Auditors make the final determination for whether a payment has an error. Furthermore, auditors may have their own incentives and biases that influence what they determine to be an error. For the OMB data, guidelines have been issued to address these concerns. These guidelines set standards (such as requiring random sampling and rigid within-program guidelines) that attempt to alleviate possible bias introduced into the data by the discretion of the auditors. Interestingly, the concerns with auditors making errors in measuring payment errors mirror those of this theory of bureaucratic errors. Errors made by auditors are likely influenced by the design of the auditing entity, its locus of implementation, the office management of the auditors (along with their incentive structures), and the auditors' own limitations, values, and motivations. The quality of the auditing data may be judged by these sets of factors, as well. This is an important concern that should be addressed when using audit data for examining payment errors.

Given the four specific aforementioned strengths, payment errors provide for an interesting arena in which to analyze the factors contributing to bureaucratic errors. While this measure should always be treated with healthy skepticism and concern for the integrity of the auditing process, it is a great starting point for testing this theory in a large-N empirical setting. In the following section, I will discuss two specific programs to which this theory can be applied, and possibly tested.

Illustrative Applications

Now that I have illustrated the different levels and the different contributing factors within each level, I would like to show how decisions at these levels may have built high payment error rates into two large US federal programs, Medicare Fee-for-Service (Parts A&B) and the Unemployment Insurance Program. These programs are chosen for discussion for three specific reasons: 1) they rank highly both in payment error rates and payment error amounts; 2) they are data-rich programs lending themselves to large-N empirical examinations (the Unemployment Insurance Program has been collecting detailed audit data for twenty years through their Benefit Accuracy Measurement program, and the Centers for Medicare & Medicaid Services collects a wide range of data on payment errors for Medicare Fee-for-Service); and 3) they are large longstanding programs that affect a large portion of the American public. In the following chapter of this dissertation I will use the Benefit Accuracy Measurement program data to begin exploring the determinants of payment error at the individual-, office, and state-level; here, however, I give brief background on each program, and provide a discussion of the factors that might contribute to relatively high payment errors for both programs. This discussion will illustrate how the theory of bureaucratic error might be used to both identify causes of error and potential improvements for the performance of major US government programs.

Medicare Fee-for-Service

Medicare Fee-for-Service encompasses parts A and B of Medicare and had annual improper payment rate of 8.5% with approximately $30 billion in improper payments. These two original parts of Medicare were established in 1965 under Title XVIII of the Social Security Act.

According to the Centers for Medicare and Medicaid Services (CMS, 2014)[8], Medicare is a health insurance program designed to serve three specific groups of individuals: 1) people age 65 or older, 2) people under age 65 with certain disabilities, and 3) people of all ages with End-Stage Renal disease. Part A of Medicare is hospital insurance, while Part B provides other medical insurance.

Part A, as a general rule, covers medical care that occurs *within* a hospital, while Part B covers necessary medical care *outside* of hospitals. Part A covers additional things such as critical access hospitals, skilled nursing facilities, and some hospice care. In general, individuals do not pay annual insurance premiums for Part A, but rather the insurance benefits are funded through a portion of federal payroll taxes. On the other hand, Part B covers things such as doctors' services, outpatient care, and other services and supplies deemed medically necessary. Part B, unlike Part A, is funded through a monthly premium. This program is thus particularly complex by design, in an attempt to be flexible enough to deal with the complex health care system.

Medicare Fee-for-Service may have the propensity for errors built into it at the design level. For example, the program pays hospitals and physicians on a model that reimburses them for a portion of the cost of services provided. This provides possible incentives for these entities to mislead the private companies that process the Medicare Fee-for-Service claims. This point also highlights the possible influence of the locus of implementation of the program. The Centers for Medicare and Medicaid Services (CMS) oversees the Medicare Fee-for-Service program, but they do not implement it on the ground. Rather, the CMS contracts with private companies to process claims, payments, call center services, clinician enrollment, and fraud investigation.

[8] http://cms.hhs.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html

Furthermore, the contracts to private companies are organized by region: a single company manages the direct implementation for an entire region of the country.

Given that private companies have been contracted to process Medicare Fee-for-Service claims, the management of these companies may differ from those of public organizations, thus possibly influencing the likelihood of errors. The rate at which payment errors occur may vary from company to company and office to office based on management factors. Medicare Fee-for-Service is also different from many other traditional programs in that the "street-level bureaucrats" administering the program are private sector workers. At the ground level these workers may have similar decision-making limitations and cognitive biases as public sector workers, but they also may have a slightly different set of values and motivators that may influence the amounts and types of errors they are likely to make.

Unemployment Insurance

The U.S. Unemployment Insurance program, known more formally as the federal-state unemployment compensation program, was created in 1935 by the passage of the Social Security Act. The purpose of this program is to provide a safety net for workers who lose their job through no fault of their own. It is a social insurance program, in which essentially all workers pay into[9] the program (through payroll taxes), and earn those benefits back in the case of approved job separation. The benefits are designed to minimize moral hazard.[10] Minimizing moral hazard is important because otherwise clients may have incentives to not work and simply live off of unemployment insurance. For example, the benefits replace only a fraction of the

[9] UI covers the majority of workers but not all. See (Department of Labor, 2013) for a more detailed explanation.
[10] Moral Hazard is an economic and finance term for situations when some party has an incentive to engage in some risky behavior because they know that they will not bear the full cost of the consequences of that decision. For unemployment insurance, minimizing moral hazard means that the program is designed in such a way to encourage people to engage in the workforce – when they are able – rather than rely on unemployment insurance benefits. If the incentives of the unemployment insurance program unintentionally encourage workers to rely on unemployment insurance, instead of working, then the program has the potential for moral hazard. The DOL has taken great care to minimize the moral hazard associated with the UI program (Department of Labor, 2013)

worker's wages, and workers must be separated from work through no fault of their own (in other words they cannot be unemployed because they quit their jobs or were fired for cause). These provisions help to keep workers from choosing unemployment over work.

The UI program is a federal-state partnership that originates from federal statute, but it is implemented and administered at the state level. According to the (Department of Labor, 2013), the federal government has four major functions with regards to UI: 1)ensure conformity and substantial compliance of state law regulations, rules, and operations with federal law; 2) determine administrative fund requirements and provide money to states for proper and efficient administration; 3) set broad overall policy for administration of the program, monitor state performance, and provide technical assistance as necessary; and 4) hold and invest all money in the unemployment trust fund until drawn down by states for the payment of compensation. The states, on the other hand, have three major functions: 1) determine operation methods and directly administer programs; 2) take claims from individuals, determine eligibility, and insure timely payment benefits to workers; and 3) determine employer liability, and assess and collect contributions. For context, the Department of Labor estimates that in 2013, 131 million workers were covered, $74.4 billion in benefits were paid, and $4.8 billion was spent on administrative costs (Department of Labor, 2013).

So, what makes this an interesting case study for the theory of bureaucratic errors, particularly payment errors? First is the sheer magnitude of the program's amount of payment errors and error rate. In 2011, the UI program had an improper payment error rate of 10.6%, and improper payments of approximately $7.6 billion (OMB, 2012). Another dataset collected by the Department of Labor's Benefit Accuracy Measurement program has a more stringent definition

of payment error, and estimates that the error rate is closer to 26%. Thus, this program makes for an ideal target because it is one of the worst offenders, in terms of both amounts and rates.

Additionally, UI is a relatively complex program; it gives a fair amount of discretion to bureaucrats, it has fairly intensive information gathering requirements, and its funding mechanism has multiple layers. Furthermore, it is an insurance program, and as such it is subject to unintended consequences, such as concerns with moral hazard. UI is also an entitlement program, which are highly political in nature. These factors, at the design level, suggest that the UI program would be likely to have a relatively high error rate.

At the locus-of-implementation level, the UI program is implemented at the state level, as opposed to the federal level. At first glance, it may be difficult to ascertain what effect this may have on the propensity for errors. However, two phenomena should result: 1) there should be variation among the states in terms of error rates; and 2) there should be significant variation in how the programs are implemented, as a result of experimentation. Thus, empirically, variation in error rates and implementation strategies should exist, and show up in the data. It might also be hypothesized that this experimentation, at the state level, should lead to lower levels of error throughout the program over time, but this does not appear to be true.

Additionally, states have a multitude of offices, and thus we should expect significant variation at the office level as well. While the currently available data do not allow researchers to look at any office-level characteristics, it is known that there is a lot of variation in payment errors at the office level (Ryu, Wenger and Wilkins, 2012; Adams and Bullock, 2013). We also know that offices which performed poorly in the previous year are likely to perform poorly in the current year (Ryu, Wenger and Wilkins, 2012; Adams and Bullock, 2013). Unfortunately, we

cannot yet trace the specific cause of these variations and trends or if different types of errors are driven by different types of factors. These are questions that I begin to address in a later chapter.

Finally, given the complexity of the UI program, the discretion available to street-level bureaucrats, the mass of required information gathering, and the range of individuals who apply, we might expect that street-level bureaucrats make systematic decisions on who gets benefits and how much they receive. We might further expect that bureaucrats develop heuristics for who is deserving of benefits and how much they deserve, given the limited time and resources allotted for each claim. Additionally, if we knew demographic information about the bureaucrats, we might expect that clients who resemble the bureaucrats might receive differential treatment.

This backstory makes the Unemployment Insurance program uniquely situated for predicting how errors are "built into" a program. The contributing factors help yield a theory explaining how we might address the concerns above, and balance them against other considerations, such as timeliness and available resources. This theory may help determine possible improvements to the current Unemployment Insurance program, or help to provide groundwork for the development of future programs.

Conclusion

The goal of this paper has been to offer a theory of bureaucratic errors, one comprehensive enough to identify the major variables that contribute to bureaucratic mistakes errors in government programs, but succinct enough to allow for the creation and empirical testing of hypotheses. The theory aims to provide insight to allow others to better design, implement, and manage new government programs, and improve upon the performance of current programs. This theory has defined errors as any deviation from an intended outcome that is mandated by either law or organizational rules. Errors are purposeful or accidental, favorable

or punitive, and signify whether they are directed at clients, other bureaucrats, the program mission, or some political power. This classification of errors allows for a prediction of how interactions and decisions made at the four levels might lead to specific types of errors.

The theory lays out four broad levels within which these errors are "built into" government programs. The four levels are program design, locus of implementation, office management, and administering by street-level bureaucrats. These strata each encompass multiple factors that directly affect the amount of errors within a program. This understanding could help legislators, policy analysts, public managers, and scholars to predict the likelihood of the amounts and types of errors that might occur within a given program. This can serve as both a diagnostic tool and a design tool.

This paper also identified a specific area in which the theory can be applied and tested: payment errors. I selected payment errors because they are clearly defined, important to the efficiency, effectiveness, and equity of government, and made at the lowest level. This allows for examination across all levels, and data are available. Research could focus across or within program variation among offices and bureaucrats. Research methods could be either quantitative or qualitative. This arena is littered with data and important questions for researchers.

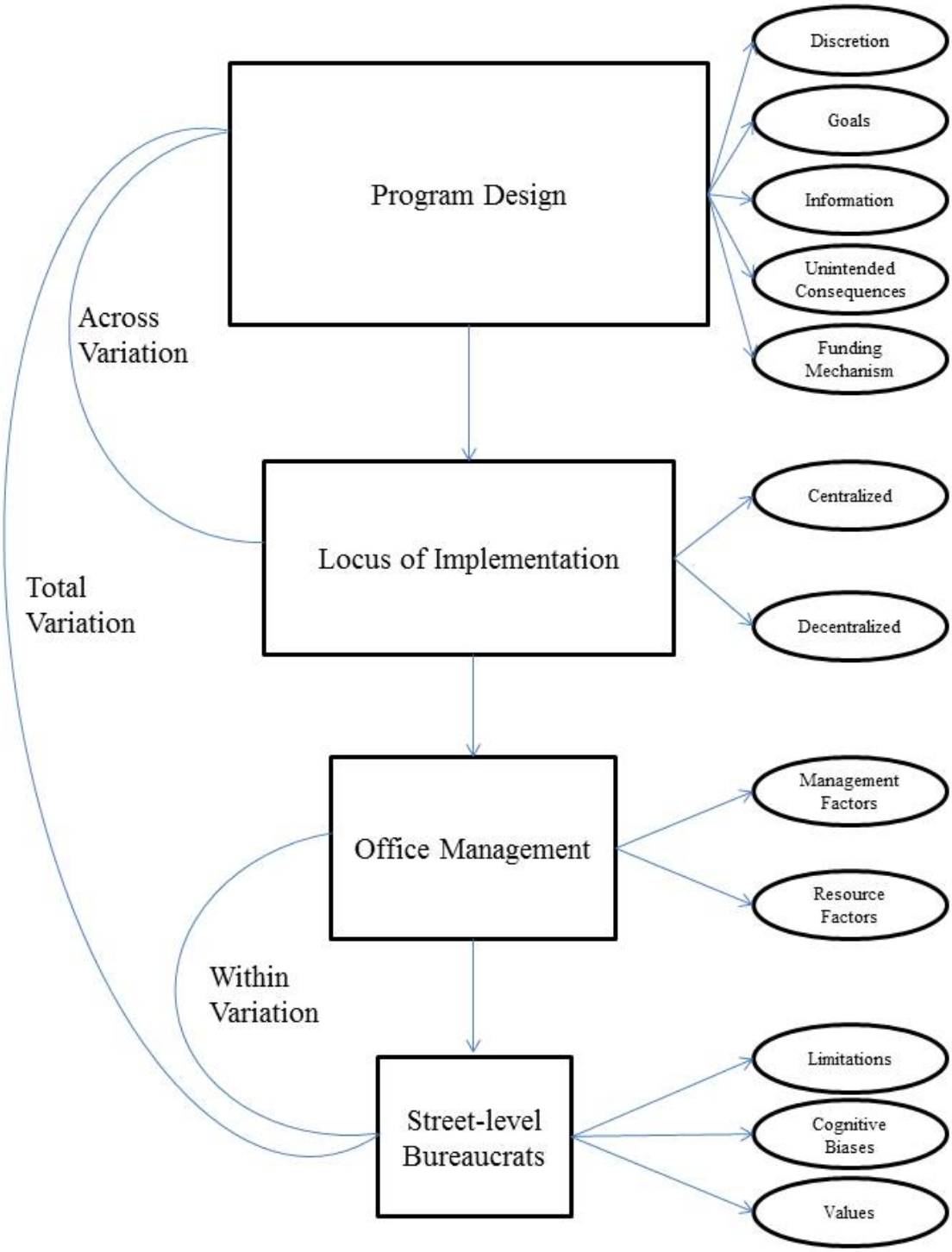
Finally, two particular programs were discussed – Medicare Fee-for-Service and Unemployment Insurance. Important decisions are made in the program design, locus of implementation, and the management of the programs; all these could affect the likelihood of payment errors within those programs. As previously mentioned, both of these programs have high payment error rates and large dollar amounts of payment errors. Understanding the workings of these programs and why they are riddled with so many errors is not only an

important public management and public policy question, but it also provides for useful cases to apply and test the theory of bureaucratic error.

This is an area of study that must be approached with some caution. Errors, particularly payment errors, are severely detrimental to the integrity and legitimacy of government, and thus, must be treated with particular diligence by public management and public policy scholars as they pursue effective, efficient, and equitable government programs and services.

Table 1.1: Theory of Bureaucratic Error: Error Categorization		
	Purposeful	**Accidental**
Favorable	Favorable, purposeful error	Favorable, accidental error
Punitive	Punitive, purposeful error	Punitive, accidental error

Figure 1: Theory of Bureaucratic Errors: The four levels



CHAPTER 2

UNEMPLOYMENT INSURANCE: AN EXAMINATION OF PAYMENT ERRORS

Introduction

This paper takes a nuanced look at the error-generating processes that drive how payment errors are made within the Federal-State Unemployment Compensation program, also known as the United States Unemployment Insurance Program (UI). In doing so, the paper draws from Bullock's (2014a) theory of bureaucratic error to identify individual claimant, office level, and state level variables that not only explain variation in payment errors across states and offices, but also develop a better understanding of how the error-generating processes may differ by error type and party responsible for the error. Furthermore, my theory of bureaucratic error provides the lens through which I examine how various levels of government programs – specifically, the levels of office management and street-level bureaucrats' decision making -- influence the likelihood of the occurrence of payment errors, and thus the error-generating process.

In examining these questions, I utilize the United States Department of Labor's Benefit Accuracy Measurement (BAM) program data. These data are doubly useful; the data identify independent variables – claimant characteristics, filing method, office caseload, and previous office error rate – that might contribute to payment errors, and allow the researcher to examine different *categories* of payment errors. The data identify both the responsible party (whose fault) and type of error (overpayment vs. underpayment). This structure lends itself to the first examination of parts of Bullock's (2014a) error typology, in which bureaucratic errors fall into one of four categories: favorable and accidental, favorable and purposeful, punitive and

accidental, or punitive and purposeful. The data cannot directly test the intent (accidental or purposeful) behind the payment errors, but the data do allow for categorization of errors by outcome (favorable vs. punitive) for one particularly interesting group, the claimants.

This paper is a first exploration of how the error-generating process may differ by error type. Specifically, it examines two main research questions: how does the error-generating process change for different types of errors (overpayments vs. underpayments)? And is the error-generating process altered based upon who is responsible for the error (agency, claimant, employer, or shared responsibility)? These two questions are analyzed using multiple levels from my theory of bureaucratic error, with a particular emphasis on office-level and street-level factors. Do street-level bureaucrats' cognitive biases and values influence the error-generating process for overpayments and underpayments differently for different types of claimants? Do office-level factors such as caseload and previous performance drive variations in errors? Do errors made by government workers resemble the same error-generating process as those errors made by employers and claimants? As will be discussed in greater detail in the data and methods section, the data allow for an examination of both the *outcome* and *responsible party* of an error, and the paper specifically examines how clients are affected by bureaucratic errors and errors that have been deemed the fault of an employer or the claimant. The theory of bureaucratic error has provided a useful categorization of errors and an understanding of how errors are built into programs (discussed below), and empirical evidence can help to trace the relationship between different types of errors – varying by outcome and responsible party – and the process by which those errors are generated.

Theory

The theory of bureaucratic error takes a broad, multi-level approach to understanding variation on errors made both across and within government programs. Bureaucratic payment errors can be one measure of the performance of government programs. The theory addresses two specific concepts – categorizing errors and identifying how errors are built into programs at four different levels – for understanding better the error-generating process of bureaucratic errors. The theory defines an error as "any deviation from an intended outcome that is mandated by either law or organizational rules" (Bullock, 2014a).

The theory of bureaucratic error includes a typology that categorizes errors along three dimensions. The first dimension is the *outcome* of the error. Errors either have a *favorable* or *punitive* outcome for some party. A favorable error is one in which an affected party receives a benefit over and above what is mandated either by law or organizational rules. An example would be an unemployment insurance recipient receiving an overpayment of benefits, which is an error favorable to the client who receives it. Punitive errors, on the other hand, withhold a benefit or provide a negative benefit to some party, such as the unemployment insurance client who receives an underpayment, or who is improperly denied. *Intent* is the second of three dimensions for categorizing bureaucratic errors. Not only can errors be categorized as either favorable or punitive (outcome), but they can also be considered as either *accidental* or *purposive* (intent). In this typology, accidental errors are those made unintentionally. There is no intent behind an accidental error; it is merely a mistake. An accidental error could be a simple missed keystroke altering the benefits of a recipient, or a rushed employee's mistaken decision, leading to an inaccurate determination of eligibility. Purposive errors, on the other hand, are made with a specific intent in mind. Picture a rogue agent who deliberately disregards the law

and organizational rules to reward a client who is not entitled to benefits, or who breaks the rules to deliberately punish a client whom they dislike. These two dimensions create four possible types of errors associated with a particular outcome and a particular intent: favorable and accidental, favorable and purposeful, punitive and accidental, or punitive and purposeful. Table 2.1 illustrates this typology using a 2x2 matrix

The third dimension is the *affected party*: who or what bears the consequences of bureaucratic errors. These parties include clients, other bureaucrats, program mission, and political principals.[11] Bureaucratic errors can accidentally punish clients, purposely favor other bureaucrats, accidentally favor the program's mission, or purposely punish a political principal. The first two dimensions (outcome and intent), as shown in Table 2.1 above, result in four separate possible outcomes for each of the affected parties. For this paper, I will be examining one affected party (the claimants) and one of the two dimensions (outcome).

To better understand the bureaucratic error-generating process, the theory of bureaucratic error identifies four specific levels at which errors can be built into government programs: program design, locus of implementation, office management, and administration by street-level bureaucrats. The first two levels (program design and locus of implementation) can explain variation in bureaucratic errors across programs, while the second set (office management and administering by street-level bureaucrats) explains variation both across and within programs. How errors could be built into government programs is illustrated in Figure 1.

A program's propensity for errors is likely to be heavily influenced by its design. The design sets the broad architecture for how a program will perform. Five distinct design factors influence the propensity for bureaucratic errors in a program: discretion, goals, information,

[11] This list, which has been included in the construction of the theory of bureaucratic error, is not exhaustive. Other groups are likely to be affected by these errors as well, at least indirectly. The above groups are the most likely to be directly affected.

unintended consequences, and funding mechanisms. Simultaneously with the program design, the decision must be made as to the locus of implementation for the program.[12] This decision (at a minimum) affects the likelihood of errors within a program, the resources available to the program, the level of experimentation in improving the program, and the influence of political actors on the program. Taken together, these two levels can help to explain variation in bureaucratic errors across programs.

The two remaining levels (and the two levels examined in this paper), office management and administration by street-level bureaucrats, can help to explain variation within and across programs. Scholars devote much attention to understanding how these two levels influence organizational performance, and thus bureaucratic errors. The factors contributing to office management can be broadly categorized into management and resources. Management factors include incentive structures, work flow systems, networking activities, and leadership capabilities; resource factors include labor supply, budget, and demand for services. The levels of office management and administration by street-level bureaucrats are the focus of this study. I will examine how both office-level variables such as office caseload, previous office error rate, and the utilization of multiple methods and how decisions made by street-level bureaucrats based upon claimant characteristics may also contribute to the error-generating process.

Unemployment Insurance

The US Unemployment Insurance Program, known more formally as the Federal-State Unemployment Compensation Program, is a particularly complex system with a wide array of moving pieces. In this section, I will discuss the basics of the system in sufficient detail to explain the purpose, basic functions, funding, and implementation of the program. Created in

[12] The theory specifically looks at the decision of choosing the locus of implementation. Policy implementation is a well-developed field of research with a wide-array of variables influencing the successful implementation of any policy. See O'Toole (2000) for an overview of several of the more prominent theories of policy implementation.

1935 by the passage of the Social Security Act, the UI Program was implemented in order to provide a safety net for workers who lose their job through no fault of their own. It is a social insurance program in which essentially all workers pay in to the program (through payroll taxes) and earn those benefits back in the case of approved job separation.

The UI program is a federal-state partnership: it is authorized by federal statute, but it is implemented and administered at the state level. According to the Department of Labor (2013), the federal government has four major functions with regard to UI: 1) ensure conformity and substantial compliance of state law regulations, rules, and operations with federal law; 2) determine administrative fund requirements and provide money to states for proper and efficient administration; 3) set broad overall policy for administration of the program, monitor state performance, and provide technical assistance as necessary; and 4) hold and invest all money in the unemployment trust fund until drawn down by states for the payment of compensation. The states, on the other hand, have three major functions: 1) make determinations about how to run the program and then directly administer the program; 2) take claims from individuals, determine eligibility, and ensure timely payment benefits to workers; and 3) determine employer liability, and assess and collect contributions. The Department of Labor estimates that in 2013, 131 million workers were covered, $74.4 billion in benefits were paid, and $4.8 billion was spent on administrative costs (Department of Labor, 2013).

The UI program is financed by both the federal government and the respective state governments. The Federal Unemployment Tax Act (FUTA) taxes cover employers at a rate of 6.0 percent on wages up to $7,000; however, states meeting federally-specified requirements receive a tax credit of up to 5.4 percent from the federal government, resulting in an effective tax rate of 0.6 percent, with a maximum of $42 per covered employee, per year. The federal tax

funds five specific expenditures: all federal and state administrative costs associated with the unemployment insurance program, the federal share of benefits paid under the federal-state Extended Unemployment Compensation Act of 1970, the loan fund from which an individual state may borrow whenever it lacks funds to pay unemployment compensation due for any month, benefits under some of the federal supplemental and emergency programs, and training services for veterans and disabled veterans.

States also collect taxes from employers to fund the UI Program. State taxes, however, vary from employer to employer based on an experience rating system.[13] Not surprisingly, these systems differ greatly across states, most notably in the way in which experience is calculated. According to the Department of Labor (2013), there are four distinct systems for calculating experience: reserve-ratio, benefit-ratio, benefit-wage-ratio, and payroll decline formulas. These differences result in an average tax rate (as a percent of total wages) ranging from 1.93 percent in Idaho to a 0.29 percent in the Virgin Islands (Department of Labor, 2013). Despite the wide array of differences, the Department of Labor (2013) notes two specific commonalities. First, in most states, three years of experience with unemployment means more than three years of coverage and contribution experience. Secondly, all of the formulas establish the relative experience of individual employers with unemployment or with benefit costs.

As with financing, coverage and benefit formulas widely vary across states. Interestingly, there are no federal standards for determining unemployment benefits; states have complete discretion in deciding qualifying requirements, benefit amounts, and duration of regular

[13] Experience rating systems adjust the rate at which an employer contributes to UI based upon the number of claims and amount of benefits paid to former employees. The more claims and benefits paid from the UI system to former employees, the higher the rate.

benefits.[14] However, states have independently developed some common guidelines on qualifying, amounts, and duration. First, in all states, the worker's qualification for compensation depends on his or her experience in the workforce during some preceding time period, called the base period. The base period wages and work experience factor into both eligibility and amount of benefits. Second, all states require that a worker has earned a specific minimum amount of wages or worked a specific minimum number of weeks (or some combination of the two) during this base period to be eligible for benefits. Third, all states require the applicant to be both able to work and actively seeking employment. Fourth, all states disqualify claimants based on their reasons for unemployment; disqualifying reasons include: voluntarily leaving work without good cause, being terminated for cause relating to work conduct, and refusing suitable work. Finally, all states employ some formula that determines the weekly benefit amount to be paid to a qualified claimant. These formulas vary considerably across states, and they have minimum and maximum weekly amounts.

These differing layers of federal and state regulation (and lack of regulation) make for a very complex system. Federal and state offices have differing roles with regard to financing the programs, and the states have wide latitude on financing (experience rating formulas), eligibility (base period, wages in base period, number of weeks worked in the base period, reasons for job separation), and benefit amounts (generosity, weekly benefit formula, and weekly benefit minimums and maximums). Furthermore, street-level bureaucrats have to make their own set of decisions on eligibility, qualification, and benefit amounts, and are allowed significant discretion

[14] There are several additional benefit programs in which the federal government provides additional resources and guidelines. These programs include: Federal-State Extended Benefits (EB),Emergency Unemployment Compensation (EUC08), Unemployment Compensation for Federal Civilian Employees (UCFE), Unemployment Compensation for Ex-Servicemembers (UCX), Disaster Unemployment Assistance (DUA), Trade Adjustment Assistance (TAA), Sell-Employment Assistance (SEA), and Short-Time Compensation (STC).

to do so. These factors render the Unemployment Insurance Program as an excellent subject for research on within-program variation on levels of bureaucratic errors.

Literature Review

To the author's knowledge, there is very little research directly examining bureaucratic errors and the error-generating process, and even less on the specific performance of the Unemployment Insurance Program. In this section, I recap the available literatures on these topics. I discuss the literatures of organizational performance and street-level bureaucrats; both of these literatures inform the hypotheses regarding the relationship between office management and administration by street-level bureaucrats and the error-generating process of a program.

Bullock's (2014a) theory of bureaucratic error (as discussed above) takes bureaucratic errors as a measure of performance and suggests payment errors as a specific measurable outcome by which government programs can be compared. Furthermore, Bullock (2014b) uses payment error rates and amounts as a measure of performance. This approach differs from many other scholars' attempts to measure performance of public organizations. Bullock concedes that this measure is not without limitations of its own, but it does provide for a useful, consistent measure for performance across a variety of settings. Scholars have used case studies, perceptual measures of performance, and other archival measures such as school pass rates and scores from the US government's performance assessment rating tool. These measures each have limitations of their own, and the use of payment errors has strengths where these other methods have weaknesses.

It is important to note that many of the studies on performance have attempted to explain variation *across* programs; however, this study examines variation *within* a program but *across* states, offices, and types of individuals. As discussed in the theory of bureaucratic error, program

implementation may be either centralized or decentralized. Unemployment insurance is implemented at the state level, and thus decentralized. This suggests that the error-generating process may not only vary in how the states implement unemployment insurance, but also in their capabilities for administering the program. Studies have found a wide array of factors at the state level that might influence government performance, including government ideology, state ideology, financial resources, and managerial capacity (Coggburn & Schneider, 2003; Andrews & Brewer, 2013; Sosin, 2012). These studies highlight state-level variation in government performance and the factors contributing to these variations.

The literature on office management is also bountiful, and represents a large portion of the body of literature known as public management. This material covers a wide array of issues and is too large to summarize for this study, but previous research has identified many factors that may affect organizational performance. These factors include design and implementation of implementation structures (Clark & Wilson, 1961; Downs, 1964; Soss, Fording, & Schram, 2011), work flow systems (Baron & Biebly,1980, 1984; Frenkel,1999), networking activities (Meier & O'Toole, 2001), management capabilities (Moynihan & Pandey, 2005), and strategy (Meier et al, 2007).

The literature on the discretion, decision making, and motivations of street-level bureaucrats is also extensive. This literature includes discussions and studies of decision-making limitations, cognitive biases, and bureaucratic values. The literature on decision-making limitations has its roots in multiple fields. Herbert Simon (1947) discussed how decision makers are likely to satisfice when making decisions constrained by time and resources. Other scholars have added to the understanding of decision-making limitations. Kahneman and Tversky's work on prospect theory and decision-making heuristics and Bazerman and Moore's (2008) work on

cognitive biases suggest that individuals (including street-level bureaucrats) are not perfectly rational decision makers, and they make decisions in biased ways. With two notable exceptions (Moynihan & Lavertu, 2012; Ryu, Wenger, and Wilkins, 2012), this literature has been largely undeveloped in the public administration field.

In contrast to the paucity of literature on decision-making limitations and biases, there is a great deal of research on how bureaucrats' values influence their decision-making processes, especially regarding who receives government benefits and how much they receive. This literature may help illuminate the error-generating processes that give rise to different types of errors. Scholars in this area have examined how clients receive differential outcomes through both active and passive representation by street-level bureaucrats (Riccucci & Meyers, 2004; Thompson, 1976; Wilkins & Kaiser, 2006). These studies also trace how bureaucrats make decisions based on who they believe to be deserving of benefits (Maynard-Moddy & Musheno, 2003; Adams & Bullock, 2013; Bullock, Wenger, & Wilkins, 2013). This literature suggests that street-level bureaucrats – through either passive representation from shared group membership or actively advocating on behalf of client groups – treat different groups of clients differently, and thus the error-generating process is likely to vary by different types of claimants.

Two specific articles examine this differential treatment by unemployment insurance agents, and they come to slightly different conclusions. Ryu, Wenger, and Wilkins (2013), using the Benefit Accuracy Program's[15] Denied Claims Accuracy dataset, find that certain demographic groups are more likely to be improperly denied than are others. Non-whites and individuals without any college education are less likely to be improperly denied. Furthermore, they find that previous office performance is a significant predictor of current performance. Ryu, Wenger, and Wilkins offer the explanation that street-level bureaucrats satisfice in their decision

[15] For a full discussion of the Benefit Accuracy Program see below in the Data and Methods section.

making. Such bureaucrats, they argue, also make decisions that suffer from cognitive limitations and psychological biases which create systematic errors. Adams and Bullock (2013) followed this with an examination of a separate dataset prepared by the Benefit Accuracy Measurement Program called the Paid Claims Accuracy dataset. In their analyses, they find that both females and blacks are more likely to receive an improper overpayment. These authors offer an explanation that builds upon Ryu, Wenger, and Wilkins' analysis: street-level bureaucrats (like all workers) suffer from cognitive biases. Adams and Bullock suggest that cognitive biases develop from social constructions of certain target populations groups, as suggested by Schneider and Ingram (1993). In their theory of social construction of target populations, Schneider and Ingram posit that groups of individuals (target populations) are categorized by society along two dimensions: positive vs. negative constructions and weak vs. strong political power. Furthermore, the authors suggested that as a result of these social constructions, different groups receive different treatment from different policies and policy tools. Bagwell and Bullock applied this typology to how street-level bureaucrats make decisions about the amount of benefits a claimant receives. For example, the authors find that both women and blacks (groups who may be viewed by the bureaucrat as positively constructed but have weak political power) are more likely to receive overpayments. They posit that these findings provide evidence for the influence social constructions may have on biasing decision-making by street-level bureaucrats. The positively constructed, weak political power groups (termed "dependents" by Schneider and Ingram) may be more likely to be the recipient of favorable errors as the bureaucrats may see this groups as both deserving and in need of the program benefits.

This paper will build upon these previous works. It uses the Benefit Accuracy Measurement's Paid Claims Accuracy data not only to examine the error-generating process of

payment errors, but also to look at how different types of errors may have different error-generating processes. In the section below, I discuss in detail the data and models to be used for this analysis.

Data & Methods

Data

To examine how the error-generating process may vary by error type, I utilize the U.S. Department of Labor's Unemployment Insurance Benefit Accuracy Measurement (BAM) Program. Given that these data are not widely used, I briefly discuss the history and objective of this program, and then I discuss how I use these data to examine both how the error-generating process changes for different types of errors (overpayments vs. underpayments), and how the error-generating process may be altered based upon who is responsible for the error (agency, claimant, employer, or shared responsibility).

The BAM program assesses the accuracy of unemployment insurance payments. According to the Department of Labor, the BAM system has three major objectives: assess the accuracy of unemployment insurance payments; assess the improvements in program accuracy and integrity over time; and encourage more efficient administration of the unemployment insurance program.

The BAM system has its roots in a National Commission on Unemployment Compensation (NCUC) study of unemployment insurance benefits in 1979 and 1980. This study examined benefits paid in six metropolitan areas and found that payment errors were occurring with much more frequency than had been previously reported. These results prompted the Department of Labor to create the Random Audit program in 1981, which would re-create the methodology of the NCUC study in five separate states. The Random Audit program ran from

1981-1984, adding more states each year; 46 states were sampled in 1984, the final year. This study confirmed the NCUC findings: payment error rates were too high. In response to these findings, in 1987 the Department of Labor declared it mandatory for states to participate in its Quality Control (QC) program. The QC program required states to sample between 500 and 2000 claims each year, according to a very specific methodology. The methodology required a representative sample that maintains statistical validity, all information to be secured through in-person contacts, timeliness of case, and publication of error rates by the state. Over the next several years, the methodology was adjusted slightly; in 1995, the program was officially changed to BAM.

BAM is thus the refined product of claims investigations processes dating back to 1979. Its random-sampling technique is very deliberate and a unique feature of assessing the unemployment insurance program. The BAM system performs detailed investigations of these randomly selected claims, thereby showing the accuracy of benefit amounts and eligibility determination. These detailed investigations include a wealth of information on the claimant, the determination made by BAM and the original UI office, and the error or lack thereof associated with the claim. The analyses are conducted using available records and contacts with the claimant, employer, and other parties. The Department of Labor promotes attempts to make the data collection as automated as possible: the sampled claims are randomly generated, the data entry is standardized using the same software across states, and the state computers link with both state mainframe computers and with the Department of Labor host computer.

These attempts to standardize the BAM system across states have met with uneven results. Individual states each have their own BAM office in which they draw samples, perform investigations, identify errors (amount of error, type of error, responsible party for the error, and

cause of error), compute error rates, and analyze data. This is important to note: despite the attempts to standardize and automate this process, it is still very likely that the decision-making processes for determinations will vary. When analyzing (as I do in this paper) variations in errors across states, it is possible that a portion of this variation can be attributed to the actual auditing process of the BAM system. The auditor themselves may be influenced by an error-generating process that is built into how BAM is designed, implemented, managed, and administered at the street-level. However, I must ignore this possible error-generating process in the remainder of the paper, since there is no audit of the BAM auditing process. For this paper, I have collected BAM data from 2002-2011 and merged the data with the State Unemployment Policy dataset. The State Unemployment Policy dataset, which covers 1990-2011, was compiled by collecting data from two primary sources made available by the Department of Labor, "The Comparison of State Unemployment Insurance Laws" and the "Unemployment Insurance Financial Data Handbook."

In examining error-generating processes and how they might vary by error type I use three dependent variables but use the same independent variables for each model. The three dependent variables are: 1) whether or not an error occurred, 2) whether the error was an overpayment or underpayment, and 3) if an error occurred, who is the responsible party (agency, employer, claimant, or some combination of these three). The following two tables give a sense of the distribution of errors. Table 2.2 shows that the total sample of claims for the ten-year period is 193,132[16]; of the total sample, 25.73% contain payment errors. The dataset contains a total of 49,698 errors. Furthermore, Table 2.3 shows that this distribution of payment errors has

[16] Cases were dropped for three reasons: 1) Data for the dependent variables for 2002 are dropped from the model as the 2002 data are used to create a lagged variable; 2) Missing data and; 3) Outliers for the wage variable (if the hourly rate is greater than $100 an hour the case was dropped from the analysis)

held fairly constant over the period. From 2003-2011 the error rate has never been higher that 28.22% and never lower that 24.7%.

Not only am I interested in the distribution of errors in general, but also the distributions of overpayments vs. underpayments, and the party responsible for the error. Table 2.4 shows the distribution of overpayments (79% of the total) and underpayments (21%). Additionally, as shown by Table 2.5, the BAM data can be coded into four possible responsible parties: agency, employer, claimant, and shared responsibility among these parties.

The independent variables for the models examine the error-generating process by individual claimant characteristics, the method by which the claimant filed for unemployment insurance, office caseload, previous office performance, and state-level variables. The individual claimant characteristics include measures of race, ethnicity, sex, age, education, and wage. The BAM data also provide information on how the claimant filed for unemployment insurance. The options for filing methods are mail (including email), telephone, employer, internet, in-person, or some non-standard method (coded as other). Table 2.6 provides the summary statistics for the individual claimant characteristics, filing methods, and office-level variables. As Table 6 shows, approximately 81% of claims are filed either by telephone or on the internet. Also, interestingly, the lagged office error rates vary widely with a mean of 25% but a standard deviation of 18%. The relationship of these independent variables with each of the dependent variables will be examined – using two different estimation techniques – to help understand how the office management and street-level bureaucrat levels from my theory of bureaucratic errors (as measured by individual claimant characteristics, filing method, office caseload, and previous performance) influence the error-generating process of both the error outcome and party responsible for the error.

Methods

As mentioned above, for the purposes of examining the error-generating process for different types of errors, I utilize three separate models. In Model 1, I examine how individual claimant characteristics, filing method, office caseload, and previous performance influence the error-generating process for *any* error. Model 1 establishes a baseline for how the independent variables may influence errors in general. Models 2 and 3 parse out this relationship by estimating the influence of the independent variables by outcome (Model 2) and responsible party (Model 3). Additionally, Model 1 uses the probit maximum likelihood estimator because the dependent variable (whether or not an error occurs) is dichotomous.

For Model 2, I examine whether the error-generating process differs from the baseline model (Model 1) when the dependent variable captures not only the occurrence of an error but additionally whether the error is a favorable or punitive error towards the claimant. Model 2 looks specifically at whether the error-generating processes (as measured by the independent variables) for overpayments and underpayments (error outcome) are different from each other and from the error-generating process for the occurrence of errors in Model 1. The dependent variable for Model 2 is coded to 0 if no error occurs, 1 if an overpayment occurs, and 2 if an underpayment occurs. To estimate this model, I use a multinomial logit. This estimator accounts for the unordered discrete nature of the dependent variable for Model 2.

In Model 3, I shift the focus from error *outcome* to error *responsibility*. Model 3 uses error responsibility as the dependent variable and analyzes, if and, how the error-generating process changes by the party or group who committed the error. It is likely that the process by which a bureaucrat commits an error is different than the process by which a claimant or employer commits an error. The variables that affect the process by which a bureaucrat commits

an error may be affected by the levels of the theory of bureaucratic errors, while claimants and employers are likely to be influenced by other motivators such as self-interest and profit maximizing. For Model 3, I estimate a five-response model for the dependent variable of error responsibility. This dependent variable – as with Model 2 – is unordered and discrete, and thus I use the multinomial logit estimator. The dependent variable is set to 0 if no error occurs, 1 if the unemployment insurance agency is solely responsible for the error, 2 if the employer is solely responsible for the error, 3 if the claimant is solely responsible, and 4 if the responsibility is shared among the groups.

As mentioned above, these three models all employ the same independent variables, which are classified into variables of interest and control variables. The independent variables of interest are those that have been discussed in detail above: individual client characteristics, filing method, office characteristics, and state characteristics. The individual client characteristics include variables on race, ethnicity, sex, age, education, and hourly wage. The filing method variables are dichotomous variables for each method (mail, telephone, employer-filed, internet, any other), and their effect on the dependent variable, as compared to filing in-person. The office characteristic variables include the office's error rate from the previous year and the number of sampled claims from each office for each year. How and why these independent variables may influence the error-generating process is hypothesized below.

Finally, the models also employ three state-level variables and dichotomous dummy variables for state and year fixed effects that serve as controls and help to capture state variation. The state-level variables are the yearly unemployment rate for each state, the number of unemployment insurance claims per capita for each state, and the number of weeks a claimant has to have worked in the base period to receive benefits. These state-level variables help to

control for variations by state and over time for changes in state policies and economic conditions. The control variables for each state capture invariant state characteristics and the control variables for time capture trends over time. The resulting standard errors from each model are clustered by state, and the data have been weighted to reflect the fact that the states do not sample the same amount of claims each week.

Hypotheses

This paper is exploratory in nature. It serves as a first examination, using large-N individual-level data, of how a variety of independent variables (individual claimant characteristics, filing methods, office-level, and state-level) may influence different categories of bureaucratic errors. It is also one of the very few studies to utilize the BAM data. Given these characteristics, it is prudent to be modest in developing hypotheses. Bullock's (2014a) theory of bureaucratic error presents an array of factors that may influence the *incidence* of government errors, but is not explicit in how these factors may contribute to different *categories* of errors. However, the theory of bureaucratic error and previous research do imply several broad hypotheses, which can be tested in this study. The hypotheses will be discussed by model.

Model 1: The Baseline Model

Model 1, the baseline model, examines the error-generating process of payment errors, generally. Ryu, Wenger, and Wilkins (2013) look at similar variables for their study on improper denials of unemployment insurance, while Adams and Bullock (2013) consider similar variables for the incidence of any error deemed the responsibility of the agency. In general, these studies have found that a host of individual-level variables and the office's previous error rate are predictors of errors. I contend that these two previous studies, the representative bureaucracy literature, the decision-making literature, and the theory of bureaucratic error would suggest that

different types of individuals will receive different types of treatment from bureaucrats. As discussed above, the error-generating process is likely to be influenced by bureaucrats' biases and values. Adams and Bullock (2013) examined how demographic groups that could be classified as dependents (positively constructed with low political power) are more likely to receive overpayments as a result of the desire of the bureaucrat to help these individuals. While weak political power can be inferred (less educated, lower wage workers, women, and minority groups) the individual bureaucrat's social construction (as either positive or negative) is likely to vary by bureaucrat and thus it is difficult to infer what *direction* (favorable vs. punitive) the error may occur, but it does suggest that claimants may be treated differently based on their individual characteristics. However, I contend that some groups are more likely to be positively construed in terms of their need or deservingness for UI. These groups include women, less-educated, and younger individuals, while attitudes towards minorities may be more mixed. Another possible explanation is that errors made by the claimants and employers in filing are actually randomly distributed, but that bureaucrats only aggressively review claims of individuals with certain characteristics. Given the nature of the data, it is difficult to rule out either of these explanations, but the explanations suggest that a variety of individual claimant characteristics may influence the error-generating process through the values, biases, and limitations of street-level bureaucrats. Thus, I hypothesize that race, ethnicity, gender, age, education, and wage will have a statistically significant effect on the occurrence of an error (H1).

Another individual-level variable, but one that is not a demographic characteristic, is the method by which the claim is filed. Filing methods have two features that are likely to contribute to the error-generating process: directness of contact with a bureaucrat and routineness of the method. The first feature suggests that the more personal contact the claimant has with the UI

agent, the more likely the agent's values and biases are to influence the error-generating process. Direct, personal contact (in-person filing) is likely to enhance the agent's sensitivity to the claimant's characteristics. The second feature is routineness. Routineness may cause errors at both extremes: most common and least common filing method. The most common filing method may have more errors as the bureaucrat may pay less attention and disengage because of a lack of interesting stimuli, while the least common filing method may have more errors because of its rarity. The bureaucrat may not as easily recognize errors when processing claims from less routine methods. This feature highlights the possible cognitive limitations of bureaucrats, and specifically UI agents. Given the two features associated with filing methods that may contribute to error, I hypothesize that the filing method will influence the error-generating process and that the filing methods will have a statistically significant effect on the occurrence of payment errors (H2).

Second, the theory of bureaucratic error suggests that office-level variables, including management and resource factors, should also influence the error-generating process. Unfortunately, the measures of office-level variables in BAM data are limited to directly observable inputs and outcomes. There are no variables available for directly assessing management and its quality. However, the observable input and outcome variables can provide suggestive evidence of the variation of performance and thus, indirectly, management. Sampled caseload is an indirect measure of resource constraints: the more sampled cases, the more claims that the office processes and thus the fewer resources the office has to devote to each individual case. Caseload as a measure of resources is likely to be muddled by each state's process for allocating financial resources to individual offices, but, irrespective of this, offices with higher caseloads are still likely to have fewer resources per individual claim. In an office with more

constrained resources, there is more likelihood of errors slipping through. Thus, I hypothesize that the offices with more sampled claims[17] are likely to be strained for resources, and thus errors are more likely to occur in such offices (H3).

The second available office-level measure is the office's error rate from the previous year. Office performance, particularly for government programs, is likely to change incrementally and slowly over time. Thus, if an office had a high error rate in the previous year, little is likely to have changed in the current year and it is likely that the previous year's error rate significantly predicts the current year's error rate. Given the wide variation in office performance (standard deviation of 18% with a mean of 25%), it can be argued that the previous error rate is an observable output of a host of management and resource factors. It is an imperfect measure that only captures one particular observable output (payment errors), but it nicely captures past office performance on this particular output and is thus an observable outcome measure that should be highly correlated with other management and resource factors. I hypothesize an office's previous error rate is likely to influence the error-generating process. Specifically, I expect that the higher the error rate for an office in the previous year, the higher the error rate in the current year (H4).

Model 2

In Model 2, I examine how the error-generating process may differ by type of error. The process that leads to overpayments is likely to differ from the process that leads to underpayments. The first distinction is that the groups that are most likely to be seen as either positively construed or deserving are more likely to receive an overpayment compared to other groups. The same pattern is not likely for underpayments. Thus I hypothesize that females, those

[17] The claims are randomly sampled at the state level. The offices with relatively more sampled claims should correspond with offices that are processing relatively more overall claims.

less-educated and younger individuals – all the groups that I believe are mostly likely to be positively construed by UI agents – will be more likely to receive overpayments but not underpayments (H5). I am less certain of the social construction of other groups. It is unclear whether bureaucrats are likely to positively or negatively construct groups such as those holding membership in a minority group or recipients of lower wages. The uncertainty of the social construction towards minority groups and recipients of lower wages leads me to hypothesize that some bureaucrats may provide favorable treatment to these groups, while others may provide punitive treatment to these groups. Thus, I hypothesize that minority groups and recipients of lower wages are more likely to receive both overpayments and underpayments (H6).

The different filing methods may also influence the error-generating process for both overpayments and underpayments. Filing methods may influence the error-generating process through either the directness of contact with a bureaucrat or routineness of the method. Direct contact with a bureaucrat may result in the bureaucrat being more sensitive to the claimant's characteristics. Thus, a bureaucrat may be more prone to favor or punish a claimant when the claimant's individual characteristics are more salient in the bureaucrat's interaction with the client (i.e. in-person filing). However, the feature that I hypothesize is the strongest driver of the relationship between filing methods and payment errors is routineness. That is, the filing methods at the extremes of occurrence (telephone claims being the most common and mail claims being the least common) are likely to similarly contribute to the error-generating process for both overpayments and underpayments. Given this reasoning, I hypothesize that the filing methods at the extremes of occurrence will be statistically significant and in the same direction for both overpayments and underpayments (H7).

Finally, for Model 2, I hypothesize that the office-level variables will similarly contribute to the error-generating processes of overpayments and underpayments. Given that the number of sampled claims is an indirect measure of resource constraints, these constraints should similarly influence both overpayments and underpayments. Thus, I hypothesize higher caseloads will have a similar effect on both overpayments and underpayments: the higher the caseload, the higher the propensity for both overpayments and underpayments (H8). The same pattern is likely to hold for the previous error rate. The higher the previous error rate, the higher the propensity for both overpayments and underpayments (H9).

Model 3

Model 3 estimates the error-generating process for four different responsible parties: UI agents, employer, claimant, and shared responsibility among these parties. This model allows me to examine how the error-generating process may differ for errors deemed the responsibility of bureaucrats compared to those errors that are the responsibility of other parties. In keeping with the theoretical explanations for Model 1, I hypothesize that the errors deemed the responsibility of UI agents will mirror those from the baseline model. Thus, I hypothesize that the individual claimant characteristics of race, ethnicity, gender, age, education, and wage will have a statistically significant effect on the occurrence of an error (H10). Also, I posit that the filing method will influence the error-generating process for the errors deemed the responsibility of the bureaucrats and that the filing methods will have a statistically significant effect on the occurrence of payment errors (H11).

Additionally, for Model 3, I again hypothesize that the offices with more sampled claims are likely to be strained for resources; thus errors are more likely to occur in such offices (H12). Finally for the error-generating process for errors deemed the responsibility of the UI agents, I

hypothesize that the higher the error rate for an office in the previous year, the more likely that an error will occur in that office in the current year (H13).

In addition to examining the error-generating process for errors committed by UI agents, Model 3 also allows me to examine the error-generating process for errors deemed the responsibility of the employer, the claimant, or involving some joint responsibility. For these error-generating processes, I develop three corresponding hypotheses. The first hypothesis is that the individual claimant characteristics should not influence the error-generating process of employer, claimant, or joint responsibility errors (H14). Errors by claimants and employers should not vary by race, ethnicity, gender, age, or education. Errors, however, may vary by wage. Low-wage workers are more likely to work in transitory jobs in which the reason for job separation may be more likely to be in dispute. The nature of these jobs may increase the likelihood of the occurrence of an error (H15). Furthermore, office-level variables should not influence the error-generating processes for errors committed by parties other than UI agents. Neither caseload nor previous office error rate should significantly predict errors committed by parties outside of those offices (H16).

Results

Model 1, the baseline model, offers insight into the error-generating process of payment errors, generally. H1 contends that, for the baseline model, claimant characteristics such as race, ethnicity, gender, age, education, and wage will significantly influence the error-generating process. As Table 2.7 shows, the results from Model 1 offer considerable support for the idea that claimant characteristics greatly influence the error-generating process. Groups that fit with Schneieder and Ingram's "dependents" categorization and who may be viewed as more deserving – groups such as women, younger workers, lower wage workers, Hispanics, and

Blacks – are all more likely to receive an error in their UI claim. Education level is much less influential. Compared to those with no college experience, only those with a graduate degree receive differential treatment. Individuals with graduate degrees are more likely to receive an error in their claim. This finding suggests that education level, a claimant characteristic that is much less easily observable, does not influence the error-generating process in the same way as more easily observable characteristics such as race, gender, and wage. This pattern of findings offers support for the notion that bureaucrats are influenced by their values and biases to provide differential treatment to different groups of individuals.

H2 suggests that the claimant's chosen filing method also influences the error-generating process. This hypothesis suggests that the features of directness of contact with bureaucrats and routineness of a particular filing method may influence the error-generating process. The hypothesis also contends that the routineness of the method may be the feature exerting the greatest influence. The results provide some evidence in support of this hypothesis. Filing in person is the comparison group for analyzing the influence of filing methods in Model 1. Fifteen percent of claimants file in-person, making this method the "middle" category in terms of the frequency with which it is utilized. Filing by mail is the least common filing method, representing approximately one percent of all claims. On the other hand, telephone claims are the most common filing method with forty-six percent of claims processed as telephone claims. These two methods – mail claims and telephone claims – are much more likely to contain errors than in-person claims. These results favor the explanation that very routine and very non-routine methods may lead to payment errors.

Moving from hypotheses about the influence of street-level bureaucratic discretion to hypotheses concerning office-level variables, H3 contends that the larger the caseload of an

office, the more constrained are the resources and the more likely an error is to occur in that office. However, the evidence does not support this hypothesis. The number of sampled claims does not significantly influence the occurrence of an error. This finding may be the result of a weak measure. The sampled claims measure is not a direct measure of resource constraint, but rather an approximate measure of the number of processed claims. It might also suggest that offices that process more claims also have more resources and thus, the measure of the number of claims processed is not a measure of resources but rather a measure of productivity by the office. Thus this finding may suggest either that this is simply a poor measure of resource constraints or that these offices are not particularly resource constrained. To parse out the explanation, a better measure of office-level resource constraints is needed.

The final hypothesis for the baseline model is H4. This hypothesis contends that changes in office performance are likely to be incremental over time and, thus the higher an office's error rate in the previous year, the more likely an error is to occur in that office in the current year. The evidence from the baseline model supports this hypothesis. Offices with higher error rates in the previous year are more likely to have errors occur in the current year. This finding suggests that the resource and management factors do not alter drastically from year to year and that change in performance is incremental over time.

While Model 1 provides evidence of the influence of both bureaucrat discretion and office-level variables on the occurrence of payment errors broadly, Model 2 examines how the error-generating process may vary for different types of payment errors, such as overpayments and underpayments. H5 suggests that the claimant groups that are most likely to be positively constructed – females, those less-educated, and younger workers – are more likely to receive overpayments, but not more likely to receive underpayments. The evidence supports this

hypothesis for females but not for the less-educated or younger individuals. Females are much more likely to receive an overpayment, but not more likely to receive an underpayment. This finding suggests that females may be seen as positively constructed and deserving by bureaucrats. While not hypothesized as such, being of Hispanic ethnicity also increases the likelihood of an overpayment but does not significantly influence the occurrence of underpayments. Interestingly, there also appears to be a punishment mechanism in place for groups who might be seen as less deserving. Those with a graduate degree are much more likely to receive an underpayment (compared to those with no college experience), but education does not appear to influence the error-generating process of overpayments. Younger workers, on the other hand, are more likely to receive both an overpayment and an underpayment. This may speak to the types of jobs held by these individuals being more transitory in nature. The same is true for lower-wage workers; they are more likely to receive both overpayments and underpayments. This finding with respect to lower-wage workers provides evidence in support of H6, which contends that the social construction or deservingness of some groups may be mixed, and thus they are more likely to receive both overpayments and underpayments. The same is true for blacks; they are more likely to receive both overpayments and underpayments, suggesting mixed social constructions and views of deservingness.

The next hypothesis, H7, suggests that filing methods should equally influence both overpayments and underpayments, and that the methods at the extreme levels of occurrence (methods that are the most and least frequent) will increase the likelihood of the occurrence of an error. However, the findings do not support this hypothesis. Instead, the results show that the filing methods at the extreme levels of occurrence – mail claims and telephone claims – influence the likelihood of an overpayment but not of an underpayment. These findings suggest

that compared to in-person claims, mail claims and telephone claims are more likely to contain overpayment errors. These findings may suggest a different explanation than the one hypothesized. There may be other factors influencing how different filing methods lead to an increased likelihood of overpayments but not of underpayments. Claimants may be more willing to lie and inflate their claim when they do not have to directly deal face-to-face with a bureaucrat. Furthermore, a bureaucrat may be more diligent in identifying errors when the claimant files in person. This could also contribute to explaining why more overpayment errors are found when the filing method involves less direct contact with a bureaucrat. While not hypothesized, evidence from Model 3 also suggests that when the error is deemed the responsibility of the claimant, this same pattern holds. That is, given an error is deemed the responsibility of a claimant, the claims filed through the mail or over the phone are more likely to contain a payment error. Incentives of employers are also highlighted in that employer-filed claims are much less likely to contain overpayments but are not a significant predictor of underpayments. With experience ratings systems in place in every state, employers have a strong incentive not to have claims that over-pay.

Hypotheses H8 and H9 contend that the influence of the office-level variables should not vary as between overpayments and underpayments, but rather that a higher caseload and higher previous error rate should increase the likelihood of both underpayments and overpayments. As with the baseline model, in Model 2 the caseload is not a significant predictor of overpayments or underpayment. Again, this finding may be a result of a weak measure of resource constraints. However, the findings do suggest that the previous office performance influences both overpayments and underpayments. This finding provides additional evidence that office performance is incremental and changes slowly over time.

The final results are from Model 3. This model examines the error-generating process by the party responsible for the payment error. To model separate error-generating processes by responsible party, the model is estimated separately for errors that are deemed the responsibility of the UI agents, the employer, and the claimant, and also for errors that are deemed the shared responsibility of multiple groups. H10 contends that individual claimant characteristics such as race, ethnicity, gender, age, education, and wage will significantly influence the error-generating process of UI agents' errors. The evidence supporting this hypothesis is mixed. Blacks, females, and lower wage workers are more likely to receive an error committed by a bureaucrat, but ethnicity, age and education do not influence the propensity of agency errors. This finding suggests that bureaucrats are influenced by race, sex, and socioeconomic status, but perhaps not by ethnicity, age, and education. As discussed earlier, this may be a result of a stronger, more uniform sense among bureaucrats of the construction and deservingness of some groups over other groups.

H11 puts forth the hypothesis that the filing method will influence the error-generating process for agency claims. There is also mixed evidence for this hypothesis. Mail claims are much more likely to contain agency error and employer-filed claims are much less likely to contain agency error, but the other filing methods are not significantly different from claims filed in person. This finding may attenuate the early finding that routineness plays a role in the occurrence of bureaucratic errors. Telephone claims (the most common filing method) are no longer significantly different from in-person claims. However, mail claims (representing one percent of claims) are still significantly different from in-person claims.

The office-level hypotheses for agency errors (H12 & H13) mirror the hypotheses from Model 1. H12 states that the number of sampled claims influences the error-generating process

for agency errors, but the evidence suggests otherwise. The sampled claims variable is not a statistically significant predictor of agency errors. H13 posits that an office's previous error rate will significantly influence the likelihood that an error occurs in the current year. The evidence from Model 3 lends support to H13. The previous year's error rate is a significant predictor of an error occurring within that office in the current year.

The final three hypotheses (H14-H16) correspond to estimations of the error-generating process for parties other than UI agents. H14 contends that the individual claimant characteristics of race, ethnicity, gender, age, and education should not influence the error-generating process for errors committed by the claimant or the employer. The evidence for this hypothesis is mixed. Ethnicity and gender do not significantly contribute to the occurrence of employer or claimant errors, but race and age are significant for both employer and claimant errors and education is significant for employer errors. Thus race influences the occurrence of an error, no matter who committed the error, but gender only influences bureaucratic errors. On the other hand, age and education do not significantly predict agency errors but do predict employer or claimant errors. These findings suggest that the error-generating process is different based upon who commits the error. Furthermore, race and age of the claimant not only influence the error-generating process of bureaucrats, but also the error-generating process for claimants and employers. H15 hypothesizes that jobs held by low-wage workers are transitory in nature and thus make for more difficult claims to process, leading to more errors. The results for Model 3 support H15. Workers receiving lower wages are more likely to have an error in their claim regardless of the party at fault for the error. Finally, H16 suggests that office-level variables should not influence the propensity of claimant errors or employer errors. However, the evidence suggests otherwise. Caseload is not significant in any of the models, but previous office error rate is significant

across all the estimations in Model 3. This suggests that even when the error is not the fault of the UI agent, some offices are better at catching employer and claimant errors before the claim is processed.

Discussion & Limitations

The goal of this paper is to provide a test for parts of Bullock's (2014a) theory of bureaucratic error. The BAM data provide an opportunity to examine the error-generating processes that are influenced by office-level variables and decisions made by street-level bureaucrats. This dataset further allows for an examination of errors generally, and according to specific types (overpayments vs. underpayments) and responsible parties (UI agents, employers, claimants, or shared responsibility). In the theory of bureaucratic error, Bullock (2014a) suggests that errors may be categorized by *intent*, *outcome*, and *responsible party*. In this study, I am able to examine separately the processes that influence different error outcomes (overpayment vs. underpayment) and errors that are the result of different responsible parties (agent, employer, claimant, and shared responsibility). The data and methods in this paper have provided evidence for several hypotheses that were created using Bullock's theory of bureaucratic error and previous research.

For this study, I find mixed support for my hypotheses. For general errors within the UI program I find support for the influence of office-level variables and street-level bureaucrats' values and biases on the error-generating process. Individual claimant characteristics such as race, ethnicity, gender, age, and hourly wage predict the occurrence of errors in a manner that is consistent with the idea that bureaucratic values and limitations influence outcomes for different types of clients. Whether through snap judgments or beliefs about who is deserving of benefits, clients are receiving differential treatment based upon their demographic characteristics. This is

further highlighted when the errors are estimated by error type. For example, Hispanics and females are much more likely to receive overpayments but are neither more nor less likely to receive underpayments. This suggests favorable treatment of these groups. Additionally, some characteristics predict a higher probability of both overpayments and underpayments. These groups include blacks, younger workers, and lower wage workers. This may suggest either that bureaucrats hold mixed views about these groups of individuals or that these types of individuals hold jobs that are more transitory in nature leading to claims that are more difficult to process. Interestingly, in Model 3 the characteristics of having a lower wage or being black influence the error-generating process for all four groups that commit errors. This suggests some program-wide biases that affect these two groups across all types of errors. For blacks, this could suggest that their racial status prompts strong responses from bureaucrats for both favorable and unfavorable treatment. For workers receiving lower wages, these findings may suggest simply that their claims are more difficult to process.

The two office-level variables offer some interesting insight across these models as well. The number of sampled claims from an office is not significant in any of the models. This likely suggests either that this measure is a poor measure of resource constraints or that those offices with larger amounts of claims have larger amounts of resources to process those claims. Particularly interesting are the findings in connection with an office's previous performance. In every model specification, the previous office error rate is a significant predictor of the occurrence of an error. This means that not only is an office's previous performance an indicator of the likelihood of a current error that is the fault of an agent; it is also a predictor of the error-generating process even for errors committed by employers and claimants. This likely suggests

that not only are offices with higher error rates more likely to commit errors, they are also more likely to miss errors committed by employers and claimants when a claim is first processed.

One final interesting and unexpected finding is the role that filing methods play in the error-generating process. I hypothesized that one important factor could be the routineness of the method, and that methods that were either very commonly used or rarely used may have more errors associated with them. The results, however, suggest a different story. Compared to filing in-person, mail claims and telephone claims are much more likely to increase the likelihood of both overpayments and claimant errors. This finding may suggest that claimants have a tendency to inflate their income when the filing method does not require them to file the claim face-to-face. It suggests that clients are more likely to lie when they are not in the direct presence of a bureaucrat. Additionally, these findings may suggest that bureaucrats are more diligent in accurately processing in-person claims, and that the bureaucrats more carefully probe the claimant for additional information when the claimant is filing the claim face-to-face. This provides some possible evidence of claimant incentives in the UI system. The same can be said for employer incentives. Employers, given the experience rating financing system, have a strong incentive not to have a claim that overpays a claimant. This is made particularly apparent in Model 2 in which employer-filed claims are much less likely to be associated with an overpayment. The same is true for employer-filed claims with errors that are the responsibility of claimants or agents.

Before I move to the conclusion, it is important to address some of the limitations of this study. As do most empirical studies, this one suffers from two important limitations, data and method. The data, while extremely useful for the purposes of this study, are collected as a result of a government program. In other words, the auditing process that collects this data may also

suffer from some of the errors associated with the unemployment insurance program more broadly. Additionally, the auditing process – while standardized across all states – is conducted separately by each individual state. This means that the processes for auditing at the state level may vary slightly, as evidenced by the wide variation of the percentage of claims with error from state to state. The Department of Labor, however, does go to great lengths to insulate the BAM program from political influences or perverse incentives. The BAM system is highly automated. Individual BAM offices throughout the country use the same procedures and reporting formats that link with the Department of Labor's host server. Each BAM state office is organizationally independent of, and not accountable to, any offices being evaluating.

In future research, I can examine the effects of other auditing systems in other programs as well; for example, the Center for Medicare and Medicaid Services contracts with private companies to audit the administration of Medicare. Comparing different auditing mechanisms could help in identifying strengths, weaknesses, and possible biases of the auditing entity and arrangement.

Another limitation of the data is highlighted in the findings section. My office-level measures are less than ideal for a nuanced understanding of how management and resource factors contribute to the error-generating process. The measures only indirectly capture office management. Office error rate captures one specific measure of performance, and sampled claims capture one type of resource availability. To better understand the underlying management processes within these organizations, I would need either survey or qualitative data to complement the BAM data. Variables at the office level, such as manager tenure, employee attitudes and pay, performance incentives, and the office's administrative budget, could be utilized to more directly measure office management and resources.

My chosen empirical methods also pose a limitation. The probit and multinomial logit were chosen as estimators, given the nature of the dependent variable and ease of interpretation. However, the independent variables of interest are hierarchical in that claims are nested within offices which are nested within different state contexts, and the models do not directly account for the nested nature of the variables. Future studies may benefit from examining the usefulness of generalized linear and mixed models in directly modelling the hierarchical variables as separate levels of analysis.

Conclusion

What drives errors in government programs? My theory of bureaucratic error begins to pursue the answer to this difficult question by identifying a multitude of factors that may contribute to the error-generating process across and within government programs. The theory suggests that one fruitful avenue for examining the concept of error is that of payment errors. Building on work by Ryu, Wenger, and Wilkins (2012) and Adams and Bullock (2013), this study takes a first look at examining the occurrence of errors by the *outcome* (overpayments vs. underpayments) and by the *responsible party* (UI agency, claimant, employer, and shared responsibility) for the error. This study finds that individual claimant characteristics and office management variables often have differing influence on the error-generating process for different types of errors. Consistent with previous studies, this study finds that characteristics of the claimant and office-level variables influence the error-generating process.

This study finds evidence to support the point that the levels of office management and street-level bureaucrat decision making influence the error-generating process of payment errors within the US Unemployment Insurance program. It also finds that the error-generating processes vary by error outcome and responsible party. The findings on individual claimant

characteristics support the idea that bureaucratic values and biases may lead to certain types of clients receiving differential treatment. The study finds that certain groups such as women and Hispanics are much more likely to receive overpayments but not underpayments, while other groups such as Blacks, younger workers, and lower wage workers are more likely to receive both overpayments and underpayments. This pattern suggests either that bureaucrats hold more mixed views on these groups or that something else about the nature of the claims submitted by these groups makes them more likely to experience payment errors. One such possibility would be the nature of the jobs these individuals are likely to be separated from.

Furthermore, this study finds that an office's previous performance is a significant predictor of an error occurring in that office during the next period, irrespective of the error type or responsible party. This finding suggests that office performance changes incrementally and that offices with high error rates not only commit more errors themselves, but also are more likely to miss errors committed by claimants and employers when the claim is first processed. This study also illustrates how the UI program design may lead to certain types of incentives for employers and claimants. Employer-filed claims are almost always less likely to experience an error and much less likely to experience an overpayment. This suggests that employers generally follow the incentive of filing accurate claims that do not result in an overpayment. Finally, the evidence also suggests that claimants may inaccurately report their work history in an attempt to receive more benefits, particularly when filing by mail or telephone. Overpayments are much more likely when the claimant does not file in-person, and the filing methods are all significant predictors of claimant errors. It may be that claimants are more likely to lie on their claim when they do not have to meet face-to-face with a bureaucrat. Alternatively, bureaucrats may also be

more diligent in correctly processing in-person claims as compared to other filing methods, which could accentuate this finding.

Future studies should continue to take a more nuanced look at not just payment errors broadly speaking, but also the type of error and the party responsible. Such examinations will aid scholars and practitioners alike with helping to find ways to minimize payment errors, and thus improve the performance of government programs that make payments to citizens. Additionally, this approach of studying payment errors can be applied to a wide range of US government programs. The study of the error-generating process, particularly for payment errors, is likely to be generalizable to a broad set of other government programs ranging from Medicare and Medicaid to the Earned Income Tax Credit program and military retirement pay programs. As discussed above, this paper highlights the important role both characteristics of claimants and office management play in the occurrence of errors. While I have identified several preliminary factors influencing errors within the UI program, further research is needed to better understand the process by which these factors combine to increase or decrease the occurrence of payment errors in government programs. Future research can utilize a variety of methods (interviews, case studies, and examination of additional empirical variables) to provide additional direct evidence of these early findings, and to help better understand the underlying processes that lead to the occurrence of bureaucratic errors.

Table 2.1: Theory of Bureaucratic Error: Error Categorization		
	Purposeful	**Accidental**
Favorable	Favorable, purposeful error	Favorable, accidental error
Punitive	Punitive, purposeful error	Punitive, accidental error

Table 2.2: Error Totals		
Error	Number	Percent
No Error	143,434	74.27%
Error	49,698	25.73%
Total	193,132	100%
Source: Benefit Accuracy Measurement data 2002-2011		

Table 2.3: Errors by Year			
Year	No Error	Error	Total
2003	16,608	5,489	22,097
	75.16%	24.84%	100%
2004	15,636	5,570	21,206
	73.73%	26.27%	100%
2005	15,689	5,299	20,988
	74.75%	25.25%	100%
2006	15,358	5,090	20,448
	75.11%	24.89%	100%
2007	15,732	5,356	21,088
	74.60%	25.40%	100%
2008	16,354	5,365	21,719
	75.30%	24.70%	100%
2009	14,879	4,897	19,776
	75.24%	24.76%	100%
2010	16,337	6,424	22,761
	71.78%	28.22%	100%
2011	16,841	6,208	23,049
	73.07%	26.93%	100%
Total	143,434	49,698	193,132
	74.27%	25.73%	100%
Source: Benefit Accuracy Measurement data 2002-2011			

Table 2.4: Overpayments vs Underpayments		
Error	Number	Percent
No Error	143,434	74.27%
Overpayment	39,370	20.39%
Underpayment	10,328	5.35%
Total	193,132	100%
Source: Benefit Accuracy Measurement data 2002-2011		

Table 2.5: Error Responsibility		
Error Responsibility	Number	Percent
No Error	143,434	74.27%
Agency Only	4,468	2.31%
Employer Only	9,930	5.14%
Claimant Only	25,648	13.28%
Shared Responsibility	9,652	5.00%
Total	193,132	100%
Source: Benefit Accuracy Measurement data 2002-2011		

Table 2.6: Summary Statistics: Individual and Office Variables

Variable	Number	Mean	SD	Min	Max
Black	193132	0.16	0.37	0	1
White	193132	0.74	0.44	0	1
Hispanic	193132	0.10	0.30	0	1
Female	193132	0.43	0.50	0	1
Age at Job Separation	193132	42.10	12.64	16	80
Hourly Wage	193132	15.90	9.60	0	100
Less than High School	193132	0.15	0.36	0	1
High School including GED	193132	0.43	0.49	0	1
Some College	193132	0.23	0.42	0	1
Associate's Degree	193132	0.06	0.24	0	1
Bachelor's Degree	193132	0.10	0.30	0	1
Graduate Degree	193132	0.03	0.17	0	1
In-Person Claim	193132	0.15	0.36	0	1
Mail Claim (including e-mail)	193132	0.01	0.11	0	1
Telephone Claim (including automated)	193132	0.46	0.50	0	1
Employer-Filed Claim	193132	0.02	0.14	0	1
Other Claim	193132	0.01	0.11	0	1
Internet Claim	193132	0.35	0.48	0	1
Lagged Office Error Mean	193132	25.88	18.15	0	100
Number of Claims	193132	96.01	117.54	1	502

Table 2.7: Results							
Variables	Model 1	Model 2		Model 3			
	Error	Over	Under	Agency	Employer	Claimant	Shared
Black	0.169**	0.301**	0.184**	0.212*	0.114*	0.349**	0.345**
White	-0.088**	-0.209**	0.047	0.006	-0.085	-0.248**	-0.053
Hispanic	0.049*	0.097*	0.012	0.133	-0.045	0.078	0.192
Female	0.046**	0.071**	0.060	0.143**	0.072	0.054	0.064
Age at Job Separation	-0.004**	-0.007**	-0.005**	-0.003	-0.006**	-0.007**	-0.010**
Some College	0.002	-0.003	0.034	0.026	0.019	-0.008	0.013
Associate's Degree	0.039	0.081	0.068	-0.016	0.180**	0.085	-0.014
Bachelor's Degree	-0.042	-0.089	-0.007	-0.024	-0.001	-0.084	-0.133*
Graduate Degree	0.070**	0.088	0.245*	0.184	0.179**	0.078	0.127
Mail Claim (including e-mail)	0.229**	0.473**	0.205	0.631*	0.053	0.402**	0.547*
Telephone Claim (including automated)	0.136*	0.288*	0.081	0.157	0.067	0.201*	0.399
Employer-Filed Claim	-0.189**	-0.448**	-0.032	-0.733*	0.214	-0.758**	-0.236
Other Claim	0.338**	0.646**	0.402**	0.763	0.328*	0.608**	0.577
Internet Claim	0.054	0.160	-0.105	0.121	-0.068	0.047	0.246
Hourly Wage	-0.025**	-0.043**	-0.060**	-0.027**	-0.080**	-0.041**	-0.035**
Lagged Office Error Mean	0.004**	0.008**	0.004**	0.006**	0.005**	0.006**	0.012**
Number of Claims (Office-level)	-0.000	-0.000	-0.001+	-0.001	0.000	0.000	-0.001
Unemployment Rate	0.023	0.050	0.013	0.103	0.024	0.016	0.106
Claims Per Capita (State-level)	7.922	9.299	42.159	9.356	37.530	-16.609	46.410
Required Weeks Worked to Qualify for Benefits	0.012*	0.030*	-0.001	0.001	0.006	0.058**	-0.040
Constant	-0.736**	-1.631**	-2.135**	-4.163**	-2.034**	-1.791**	- 3.483**
Observations	170066	170066		170066			
Significance Levels : * p<0.05 ** p<0.01							

CHAPTER 3

PERFORMANCE IN THE PUBLIC SECTOR: A LOOK AT FEDERAL IMPROPER PAYMENTS

Introduction

This paper offers an initial, partial test of my theory of bureaucratic errors. In this study, federal, program-level, improper payments data are used to examine which factors predict performance of the program. I use the Office of Management and Budget's (OMB) Improper payment dataset of approximately 100 federal programs from 2004-2011, which provides improper payment amounts and improper payment rates that I use as measures of performance. These programs span eighteen agencies and cover a wide variety of programs from public insurance programs to early education programs and from the United States Agency for International Development (USAID) programs to pension and other retirement programs. In 2011 alone (the most recent year for which data are available), of the approximately $2.4 trillion paid through these programs, $115 billion were improper with $89 billion of those payments being overpayments.[18] The average error rate of these programs from 2004-2011 is approximately 4.4%, but the range of the programs' error rate in a given year is from an error rate of zero to an error rate of 62%. Furthermore improper payment amounts for a program in a given year range from $0 to $30.8 billion. These numbers suggest wide variation in the performance of these programs with respect to improper payment rates and amounts.

[18] These numbers suggest that 77% of improper payments are overpayments. While this is strictly true, $15 billion in improper payments (from the Earned Income Tax Credit) are not classified as either overpayments or underpayments. If these numbers are removed from the calculation then overpayments account for almost 90% of all payment errors.

Payment errors can take many forms, depending on the program. One high-error program is the Earned Income Tax Credit (EITC). The EITC had a payment error rate of 23.5% in 2011 and an average payment error rate of 25.19% from 2004-2011. The EITC is designed to provide a tax break to lower income Americans who work. Qualification for EITC is a function of income and the number of dependents a household has. Furthermore, the credit phases in and phases out slowly so as to reduce disincentives to work. Unfortunately, the eligibility requirements (income and number of dependents) and the variability in the credit amount leave the program susceptible to payment errors. For example, EITC beneficiaries are not pre-certified as with other benefit programs, such as Unemployment Insurance and Medicare.[19] Additionally, a number of other contributing factors add to the higher error rates of EITC, including high turnover in those who receive EITC, errors made by tax preparers, and fraud. Other programs, however, have different processes that may lead to high error rates as well.

Another example of a high error rate program is Medicare Fee-for-Service (Parts A & B). The goal of Medicare is to provide health insurance for people who are age 65 and older, regardless of their medical history or income level. Currently, Medicare is composed of four parts (A,B,C, & D). Medicare Part A covers inpatient hospital care. Medicare Part B is an insurance program that covers outpatient medical services and is the traditional method by which Medicare offers insurance. Taken together, Part A and Part B constitute the program known as Medicare Fee-for-Service. Medicare Fee-for-Service had an error rate of 8.6% in 2011 with $28.8 billion in improper payments. The improper payments for Medicare Fee-for-Service arise from a different set of factors than do those for the EITC. For example, the primary causes of payment error in Medicare Fee-for-Service include administrative and documentation errors,

[19] In other words, EITC claimants receive the credit based upon their tax return, whereas eligibility for other major government programs is determined by a bureaucrat through some certification process.

along with authentication and medical necessity determination errors. Medicare Fee-for-Service and EITC accounted for roughly $44 billion in improper payment in 2011. While different underlying processes help contribute to these payment errors, this study and my theory of bureaucratic error seek to find common predictors that can help to explain the variation in errors in a wide range of government programs.

In this paper, I use improper payments to measure performance; this conceptualization has not yet been tested in the public management literature. This approach will allow for an examination of change in payment errors over time by program, organizational-level characteristics that might predict error rates, and employee perceptions, to explain the variance in payment errors. This study offers an exploratory test of how the different levels (program design, locus of implementation, office management, and street-level bureaucrats) identified in my theory may influence the prevalence of payment errors and thus performance more broadly for these federal programs.

Why the emphasis on payment errors? Generally, researchers have taken one of four paths for examining the performance of government programs. The first of these approaches is the case study (Carlson, Cowen, & Fleming, 2013; Eisenhardt, 1989; Gormley, 1996; Schneider, 2004), a holistic approach to the performance of a program. Case studies examine a multitude of variables within one program and delve into the minutiae that help to better explain the causal story of a program's performance. While this strategy is very useful for examining a single program in great detail, it is difficult to generalize the results to other programs (Eisenhardt, 1989).

One can also use perceptual data to examine government performance. Perceptual data have become popular in the public management literature (Kenneth J. Meier & O'Toole Jr,

2013), because they allow for ease of comparison among different programs and organizations. This method, however, has serious weaknesses. Researchers usually utilize data from the same source for both the perceptions of performance and the independent variables of interest. This can introduce common source bias, a serious methodological concern that can lead to false positive findings (Favero & Bullock, 2014; Kenneth J. Meier & O'Toole Jr, 2013). This approach also raises the problematic question of whose perception of performance is to be measured (line-workers, middle-managers, department heads, or a weighted average of all three)? Supervisors and supervisees may have completely different perceptions of not only what performance is, but also how well their organization or program is performing, based upon their respective definitions of performance (Provan & Milward, 1995). These concerns could introduce large amounts of error into the measure of performance, yielding suspicious results.

The third path for measuring performance is an archival measure that is specific to a particular type of program. For example, in testing out their managerial networking model, Meier & O'Toole have often used pass rates on a Texas standardized test as a measure of organizational performance at the school district level (K. J Meier & O'Toole, 2003, 2005; O'Toole & Meier, 2004). Arguably, a great strength of this approach is construct validity of performance. For example, pass rates are often considered to be a good metric for examining school performance. Unfortunately, this approach is fairly limited in its generalizability (i.e. pass rates are a measure specific to schools). On the positive side, these results should tell us something about managerial influence in school districts in other states. On the negative side, the measure of pass rates is much more difficult to extrapolate to other types of organizations. Managerial behaviors that help to make public schools perform well may not be as helpful for managing an unemployment insurance office.

The fourth and final common approach for examining performance is the US government's Performance Assessment Rating Tool (PART). PART scores were developed as a performance management measure by the OMB during the George W. Bush administration. The measure was designed to assess four separate types of performance: program purpose and design, strategic planning, program management, and program results. The scores on each of these categories were then combined to create an overall program evaluation. These scores were collected once for each federal program from 2002-2008 (Gallo & Lewis, 2012). This measure functions across many different types of U.S. federal programs, and while limited to this particular sample, it would still include programs operating within a wide range of managerial or policy settings. Unfortunately, researchers have found this measure to be riddled with a suspicious amount of measurement error, and generally less than ideal (Heinrich, 2012; Radin, 2006; Gilmour & Lewis, 2006) .

The shortcomings above help support the selection of error rate and error amounts as superior options for studying the performance of public programs. This method has weaknesses of its own, which will be addressed later, but its advantages are significant. The minimization of errors method will allow researchers to generalize findings to similar types of programs, address a type of performance vital to retaining public legitimacy, and deal with questions answerable with available data. This approach also allows researchers to examine the question of performance more broadly, by recasting performance as a question that involves design, implementation, office management, and bureaucratic attitudes. I offer four dependent variables for study: error dollar amounts, logged error dollar amounts, error rates, and perceptions of performance. The measure of error dollar amounts is a measure of the mounts of payments that are errors. This measure allows for an examination of the variance in the total dollar amounts of

payment errors. However, given the wide variation in the size of programs – in terms of the amount of payments to the public – I also examine errors in logarithmic form and as a rate. The logarithmic form helps to control for the skewed nature of the data given the wide variation in payment amounts. The error rate looks at errors as a proportion of total payments rather than just examining absolute dollar amounts. The error rates standardize the measure across widely different sizes of programs. The fourth and final measure is a perceptual measure. This measure is used to highlight the differences in findings between archival measures and perceptual measures. As I will show, there are very significant differences between the results of the payment error models and the perceptual model. In explaining these findings I offer a discussion of the importance of common source bias in using perceptual measures for performance.[20] The theory behind the approach of error (particularly payment error) minimization as a measure of performance follows, broken down across the four levels at which errors are built into programs (program design, locus of implementation, office management, and administering by street-level bureaucrats).

Theory

The field of public administration, from its earliest roots, has been concerned with how effectively and efficiently to implement government programs (Wilson, 1887); this means that the field has always been concerned with understanding and evaluating the performance of public organizations and public programs. Unfortunately, this study has often progressed in a very piecemeal fashion, with little general theory. The theory of bureaucratic error is designed to remedy this shortcoming. This theory, by design, incorporates several piecemeal approaches into

[20] Common source bias is a statistical concern that arises when a multivariate model uses data from the same source for perceptual dependent and independent variables. The concern is that the common source (for example, using questions from the same survey as dependent and independent variables) explains a significant amount of the correlation between the independent and dependent variables rather than the attitudes and perceptions themselves.

a broader theory, which focuses specifically on errors as a measure of performance. All types of organizations can aspire to error minimization, and the consequent gains to performance.

I begin by addressing two foundational questions for this line of thought. What is the most helpful way to categorize errors? How, and at what levels, are errors built into programs? What follows is a quick recap of the theory and a discussion of the additional literature informing the hypotheses to be tested.

An error is any deviation from an intended outcome that is mandated by either law or organizational rules (Bullock, 2014). This definition is intentionally broad, in an attempt to understand a wide array of different types of errors, within three important categorizations: *purposive vs. accidental, favorable vs. punitive,* and the *affected party.* One useful way for thinking about these categorizations is to use a 2X2 matrix.

As Table 3.1 demonstrates, any error can be thought of as either being purposeful *or* accidental, and either favorable *or* punitive. A purposeful error is deliberate and intentional, and made to directly favor or punish some party. Accidental errors also directly favor or punish some party, but different from purposeful errors in their lack of *intent*.

These errors can also be categorized by *outcome.* Is the error favorable or punitive to some group or individual? An error is favorable when the affected party receives some positive treatment over and above what is mandated either by law or organizational rules; an error is punitive if the affected party receives some negative treatment or lack or treatment that is mandated either by law or organizational rule. These errors may also affect a wide array of parties. The theory specifically addresses how four particular groups directly affected by bureaucratic errors, including clients, other bureaucrats, program mission, and political principals.

How, and at what levels, are errors built into government programs? The four levels identified in my theory through which errors may be built into programs are program design, locus of implementation, office management, and administering by street-level bureaucrats. Factors at each of these levels may lead to more or fewer errors in government programs.

At the *program design* level there are, at a minimum, five factors to be considered: discretion, goals, information, unintended consequences, and funding mechanism. The second level, *locus of implementation*, involves the tradeoffs between centralized and decentralized implementation. Next comes *office management*, within which the subsets of management factors and resource factors can affect propensity for error. Finally, the fourth level is *the administration of programs by street-level bureaucrats*. Street-level bureaucrats are influenced by at least three sets of variables: decision-making limitations, cognitive biases, and bureaucratic values.

This study seeks to examine the variation in bureaucratic errors across programs by focusing on payment errors.[21] This is a first step in exploring how factors at the different levels might affect the propensity and amount of payment errors within a U.S. federal program. This study will not examine individual payment errors, and thus will be unable to make claims about different types of errors. The OMB's improper payments dataset provides program-level data of improper payments made to clients (recipients of benefits), and captures both over- and under-payments. Unfortunately, the dataset does not provide individual-level (claim-level) data for improper payment errors, so the analysis will be limited to the program-level of analysis. This study will examine how factors within program design, locus of implementation, office

[21] Variation in error rates across programs is only one way in which to utilize this theory. Across-program error variation focuses on the levels of program design and locus of implementation, but the theory can also be used to analyze within-program error variation. The levels of office management and administration by street-level bureaucrats easily lend themselves to examining within program error variation.

management, and administration by street-level bureaucrats affect the improper payment error rates and amounts of improper payments.

Literature Review

In seeking the predictors of organizational performance, public administration scholars have examined several independent variables, including goal ambiguity (Chun & Rainey, 2005), program types (Jung & Rainey, 2011), organizational goal characteristics (Jung & Rainey, 2009), management (Boyne et al., 2011; Cho & Ringquist, 2011; K. J Meier & O'Toole, 2002, 2003; O'Toole & Meier, 1999, 2003; Walker & Andrews, 2013), Government Performance and Results Act (GPRA) processes and Performance Assessment Rating Tool (PART) scores, and size of the program (Jung & Rainey, 2009). In addition, federal survey responses to multiple years of the Federal Viewpoint Survey capture mean and variance of organizational-level employee attitudes that could influence performance (Bertelli, Mason, Connolly, & Gastwirth, 2013; Sabharwal, 2013).

Additionally, (Jung & Rainey, 2009) attempt to estimate organizational performance, as measured by PART scores, as a function of assessment year, program type effects, and program goal ambiguity. (Moynihan & Lavertu, 2012) look at predicting the use of performance information as a function of PART scores, GPRA, and a set of individual characteristics. Additionally, (Chun & Rainey, 2005) measure organizational effectiveness (an aspect of performance) using four separate employee perceptual measures: managerial performance, customer service orientation, productivity, and work quality. These measures are then estimated as a function of individual- and organizational-level characteristics such as organizational size, organizational age, policy problem complexity, competing demands, financial publicness, mission ambiguity, directive ambiguity, evaluative ambiguity, and priority ambiguity.

These studies help explain perceptions of organizational performance and PART scores as a measure of performance. However, as previously mentioned, the first set of results may exhibit common source bias, as both the dependent and independent variables of interest are measured with the same survey tool (Huselid, 1995; Kenneth J. Meier & O'Toole Jr, 2013; Kenneth J. Meier & O'Toole, 2013). The second set uses PART scores as a reliable indicator of performance, although scholars have also found these to be problematic (Radin, 2006). My approach avoids these concerns, and allows for an examination of changes in error rates over time by program, organizational-level characteristics, and employee perceptions.

Data & Methods

Data

In 2002, the 107th Congress passed the "Improper Payments Information Act of 2002." This short act (only two pages long) charged federal agency heads with: 1) estimating the annual amount of improper payments made in their programs; and 2) submitting the estimates to Congress before March 31 of the following applicable year, with all agencies using the same method of reporting, as determined by the Director of the OMB (see Appendix B for the source document). In 2010, the 111th Congress passed an amendment to the Improper Payments Information Act of 2002. This amended act, called the "Improper Payments Elimination and Recovery Act of 2010," was instated "to amend the Improper Payments Information Act of 2002 (31 U.S.C. 3321 note) in order *to prevent the loss of billions in taxpayer dollars"* (emphasis mine). This amendment is slightly longer than the initial version (13 pages), and provides for more specific guidance on how agencies are to comply with the legislation. This legislation also

places a renewed focus on catching improper payments, identifying ways to eliminate them, and developing systems for recovery.[22] Table 3.2 displays the summary statistics for the OMB data.

The OMB maintains this dataset and makes it publicly available on their website.[23] It examines approximately 100 federal programs from 2004-2011, and provides agency name, program name, fiscal year, active/inactive status, total amount of program outlays, total amount of program outlays that are improper, total amount of outlays that are overpayments, total amount of outlays that are underpayments, three-year projections of outlays, and three-year projections of improper outlays. I merged it with another data set that includes mean responses of attitudinal questions, by agency, from the U.S. Federal Viewpoint Survey. These approximately 100 programs are identified by the OMB as susceptible to significant improper payments. These programs are housed in 18 agencies and include programs from HeadStart to Medicaid and from School Lunch to Military Retirement Benefits.

Method

The multivariate method used for this line of inquiry is the generalized least squares random-effects model. Three of the dependent variables of interest (improper payment rates, improper payment dollar amounts, and improper payment amounts in logged form) are continuous variables, and thus generalized least squares is appropriate. The fourth dependent variable is an averaged response that ranges from 1-5 and thus is not continuous (rather it is censored at the values of one and five), but the same estimator is used, for ease of comparison with the previous three dependent variables.[24] The data are an unbalanced panel, and the standard errors are clustered by agency. Four separate models are estimated. The first model

[22] Legislation updating the original Improper Payments Information Act of 2002 was passed in 2010 and 2012. The general purpose of these statutes was to further clarify and specify requirements put forth in the original legislation.
[23] http://www.whitehouse.gov/omb/financial/improper_payment_dataset/
[24] Model 4 was also estimated using a tobit with one as the lower limit and five as the upper limit. The results were not substantively different

uses the improper payment rate as the dependent variable; the second model uses the improper payment amount as the dependent variable. In the third model, the logarithmic form of the improper payment amount is used as the dependent variable to address the skewed nature of the improper payment amount. Finally, in a fourth model, a perceptual measure of performance is used as a comparison point to the archival measure of improper payments.

The same independent variables are utilized consistently throughout all four models. The first independent variable in each model is a lagged performance variable; the second is the total amount of outlays for a program in each year. The remaining eighteen variables are averaged perceptual attitudes at the agency level. These items represent a sample of questions that appear in the Federal Employee Viewpoint Surveys (FEVS) from 2004-2012.

The FEVS has been administered by the Office of Personnel (OPM) since 2002. First established as the Federal Human Capital Survey, it has been conducted in 2002, 2004, 2006, 2008, 2010, 2011, 2012, and 2013. In 2012, the most recent year for which the data are analyzed for this study, 1.6 million federal employees were surveyed, with over 687,000 responses. The survey has three goals: 1) to provide general indicators of how well the Federal Government is running its human resources management systems; 2) to serve as a tool for OPM to assess individual agencies and their progress on strategic management of human capital; and 3) to give senior managers critical information to answer a vital question: what can I do to make my agency work better?[25] In 2012, the FEVS continued to improve its statistical usefulness by attempting to reach every full- or part-time, permanent, civilian government employee.[26] For the 2012 survey, 82 agencies, consisting of 37 departments or large agencies and 45 small or independent

[25] http://www.fedview.opm.gov/2012/What/
[26] http://www.fedview.opm.gov/2012/

agencies, took part in the survey; these agencies represent 97 percent of the employees within the executive branch.[27]

Hypotheses

The following hypotheses test the influence of factors at several levels that are identified by my theory of bureaucratic error. In this study, each hypothesis is addressing *across* program variation as opposed to *within* program variation. H1 and H2 focus on factors associated with program design. H1 examines the inertial nature of programs by examining how given a certain design, programs are unlikely to change in their error rates and amount from year to year by hypothesizing that a program's previous performance is likely to strongly predict current performance. H2 hypothesizes how the size of program is likely to influence its propensity for errors, again a factor associated with program design. H3, H4, and H5, on the other hand examine across program variation in the average attitudes of employees within these programs. Bureaucratic attitudes are factors associated with both the office management and street-level bureaucrat levels. These hypotheses will examine the relationship of perceptions of job advancement, job satisfaction, and pay satisfaction with payment errors. These variations may be explained in part by management factors such as management quality and the influence it has on employee attitudes, and also by street-level bureaucrat factors which highlight the connection between bureaucrats' attitudes and motivations with errors.

One predictor of current payment errors or current performance should be the program's previous performance (Ryu, Wenger, & Wilkins, 2012). Government programs are often inertial by design and thus changes in these programs should be incremental over time. Other than large external shocks, these government programs are unlikely to change greatly from year to year. If a

[27] http://www.fedview.opm.gov/2012/What/

program has a high error rate or a large amount of errors in a previous year, they are more likely to have a high error rate or large amount of errors in the current year (H1).

Larger program budgets are an indicator of overall program size. Larger programs are likely to be more extensive in their scope and have more decision makers who influence the program outcomes. These characteristics expose programs to more opportunities for errors. Additionally, larger programs may have a larger capacity for oversight and management, decreasing the likelihood of errors. Thus, I hypothesize that the program budget will affect the improper payment amount, but am uncertain as to the direction of this relationship (H2).

If employees within an organization perceive that they are able to advance to better jobs within their organization, they are more likely to work towards these advancements. This should lead to improved performance and decreased errors within the organization. Thus, I hypothesize that the higher the perceptions of opportunity within an organization, the lower the improper payment rate and dollar amount (H3).

In previous studies (Judge, Thoresen, Bono, & Patton, 2001; Kim, 2005; Vandenabeele, 2009), perceptions of job satisfaction have been found to be related to perceptions of performance. Employees who are satisfied with their job are more likely to be more productive and perform at higher levels. Thus, I hypothesize that the higher the perception of job satisfaction, the lower the improper payment rate and amount (H4).

Pay satisfaction is an important factor in motivating employees (Brewer, 2005; Jung, 2014; Lee & Whitford, 2008; Yang & Kassekert, 2010). If employees are not satisfied with their pay, they are less likely to perform at their highest capable levels. Thus, I hypothesize that the higher the perception of pay satisfaction, the lower the improper payment rate and amount (H5).

Results

In discussing the results, Models 1-3 will be analyzed together, as they each use the archival measures of improper payments errors. Model 4 will be handled separately, with an emphasis on its differences from the first three. Table 3.3 displays the results

The first interesting finding is that all three models have high r-squared's: .6128, .9596, and .8227, respectively. This suggests that the model explains a sizeable amount of the variance in the dependent variable. Models 1-3 provide strong evidence for H1, in both statistical significance and magnitude. In Model 1, a percentage point increase in the previous year's error rate is associated with a .89 percentage point increase in the current year. In model 2, a $1000 increase in the previous year's improper payment amount is associated with a $954 increase in the current year. Finally, in Model 3, a one percent increase in the amount of improper payments in the previous year is associated with a .85 percent increase in the amount of improper payment in the current year. Taken together, these three models strongly suggest that the program's previous performance strongly predicts current performance.

While all three models support H1, the other four hypotheses do not fare so well. H2 hypothesizes that variations in the size of the program are likely to affect the performance of a program, but this hypothesis does not hold across the three models. The amount of program outlays is not significant in either Model 1 or 2; it is marginally ($p<.10$) significant in model 3, but not substantively. These findings suggest that program size, as measured by the amount of outlays of a program, is not a predictor of performance, controlling for the other variables in the model.

Models 1-3 also do not lend support to H3. The survey item corresponding to advancement opportunities is not significant in any of the models. This is also true of H4; the

item for job satisfaction is not significant in Models 1-3. Thus, these models strongly suggest that agency average responses to attitudes concerning both advancement and job satisfaction are not predictive of program payment error rates or program improper payment amounts.

The final hypothesis, H5, also finds little to no support in Models 1-3. The item representing pay satisfaction is not significant in Models 2 or 3 and is only marginally ($p<.10$) significant in Model 1. These findings suggest that satisfaction with pay is not a strong predictor of performance while controlling for the other variables in the model. Finally, it is also interesting to note that almost none of the other perceptual variables are predictive of performance in Models 1-3. This is particularly surprising given the longstanding organizational research that suggests that these attitudes do indeed play a large role in performance.

Model 4, however, paints a different picture. As with Models 1-3, Model 4 offers strong support for H1. The previous year's performance (program error rate) is a positive and significant predictor of current performance. This suggests that employees' perceptions of programs' past performance are a strong predictor of their perceptions of the program's current performance. Unlike Models 1-3, Model 4 finds some support for H2. The total dollar amount of program outlays has a negative and significant effect (at the $p<0.05$ level) on the perception of performance. This suggests that the larger the amount of outlays for a given program, the lower the perceived performance. For H3, Model 4 finds the perception of within- organization opportunities is related to perceptions of performance, but in the opposite direction to the hypothesis. Model 4 suggests that the higher the perceptions of within-organization opportunities, the lower the perceived performance. As with Models 1-3, Model 4 does not find evidence to support the hypothesis that job satisfaction is related to performance (H4). Additionally, with regards to H5, Model 4 finds support that suggests pay satisfaction is a

positive and significant predictor of perceptions of performance. Finally, unlike Models 1-3, Model 4 finds that many of the perception variables are significant predictors of the perceived performance.

Discussion & Limitations

This partial test of my theory of bureaucratic errors has yielded some cautious initial findings. The programs analyzed are strongly influenced by the previous year's performance. The higher the error rate of the program in the previous year; the higher the error rate for the program is likely to be in the current year. Second, the size of the program's outlays does not appear to influence the dollar amount of errors or error rates. Finally, a wide variety of employee attitudes do not appear to be related to the dollar amount of payment errors or error rates. These conclusions have important ramifications for the field. First, the lack of influence of employee attitudes on performance runs directly contrary to the findings of other scholars. While important, employee attitudes may not be as important as other factors to program variation in performance. Second, large or small programs (in terms of dollar amounts of payments) are not necessarily poor or great performers. Large programs are not destined to be riddled with large amounts of improper payments; smaller programs are not consistently more error-free. Third, previous performance is a very significant and substantive predictor of current performance. This suggests that programs that were poor performers in the previous year are very unlikely to change their performance dramatically in the next year; furthermore, programs that were great performers in the previous year are likely to remain strong performers in the current year.

What do these findings suggest? What does it mean that a program's previous performance is the only consistent predictor of current performance? More information and research are needed to answer these questions completely, but some cautious initial inferences

can be made. Performance of these programs changes very slowly. Additionally, the most likely targets for reducing errors *across* these programs would seem to be *program design* and *locus of implementation*. These two levels contain multiple factors that may help to explain the wide variation in errors and error rates in the government programs. For program design, the level and type of discretion granted to bureaucrats, the nature of the goals of the program, the intensity of information gathering and the information asymmetry, the unintended consequences of the design, and the funding mechanism each may play a significant role in explaining the variation in errors in these programs. The finding regarding the influence of a program's previous performance on current performance highlights the need for a renewed focus on the relationship between the nuts and bolts of policies and the amount of payment errors likely to be generated in a program. Furthermore, the implementation arrangement likely influences the variation in error in these programs. As discussed in my theory of bureaucratic error, the decision of whether to centralize or decentralize the implementation of a program is likely to influence the amount and rate of errors of that program. Still, additional studies are needed to confirm or disconfirm these ideas.

Table 3.4 displays some of the best performers (those with a less than 1% error rate) and worst performers (those with an error rate of 10% or higher). Examining these programs through the lens of the four levels put forth in my theory of bureaucratic error may help to explain the variation across these programs, and particularly to explain why some programs have payment error rates at or near zero and other programs have error rates over twenty percent. The evidence from this study provides preliminary evidence suggesting that when examining error rates *across* programs, factors at the program design level may matter greatly for the variation in payment errors while differences in office management and bureaucratic attitudes may matter less. The

measures in this study for program design factors, office management factors, and bureaucratic attitudes, however are indirect measures of the factors that may be driving the variation in errors; future studies will need to develop more direct measures of the factors at each of the four levels (program design, locus of implementation, office management, and street-level bureaucrats) to provide stronger evidence for these initial findings.

Within-program variation for payment errors in the U.S. Unemployment Insurance (UI) program has been examined by Ryu et al. (2012) and Adams & Bullock (2013). The U.S. UI program is an excellent government program for the study of payment errors, since the program has a relatively higher error rate, is implemented differently by each state, and places a heavy burden on street-level bureaucrats for the degree of information that they must collect in making determinations about eligibility and benefit amounts. While these studies do not examine program-level design characteristics, they do highlight factors which influence *within*-program payment error variation, such as state administration, office management, and bureaucratic values. The third paper in this dissertation undertakes a similar analysis. The research thus far suggests that while previous performance, program design, and program implementation may matter greatly when explaining variation *across* programs, other factors such as administration, office management, and bureaucratic values directly influence variations *within* programs.

This paper also highlights the importance of how studies define and operationalize performance; this study demonstrates that if a subjective measure is substituted in for the dependent variable of performance, as opposed any to any of the archival improper payment measures, then the findings greatly differ. When the perceived measure of performance is used, many of the perceptual variables become significant predictors of performance. Not only is previously perceived performance an important predictor of currently perceived performance,

but also *half* of the perceptual variables are significant predictors of current performance. While it is not, strictly speaking, appropriate to directly compare archival models with perceptual models, it is important to note that once a common source method of attitudinal variables is used on both sides of the multivariate equation, a wide variety of attitudinal variables begin predicting performance. It is likely that Model 4 is plagued with "false positives." These potential false positives are likely due to correlated measurement error between the dependent variable and attitudinal variables in Model 4. Much empirical research has been done that suggests when attitudinal variables from the same survey are analyzed in a multivariate model the variables of interest may be correlated, but the correlated measurement error may also be the driving force behind the observed correlation (Meier and O'Toole 2013; Podsakoff, MacKenzie, and Podsakoff, 2012). For example, it could be that an individual's true pay satisfaction and their true perception of an organization's performance are correlated, but it may also be that the correlation is actually capturing correlated measurement error. Favero and Bullock (2014) highlight that in multivariate models in which the dependent and independent variables are both attitudinal and measured from the same source, the positive findings could indeed be false positives, rooted in common method bias. Findings using common source methods that suggest employee attitudes play such a large role in performance may need to be treated with caution.

Conclusion

This paper has aimed to offer an alternative measure for testing performance, in the form of improper payment errors, and provide a first test for Bullock's (2014) theory of bureaucratic error. Two main findings arise. First, the choice of dependent variables (as between a perceptual and archival measure) has a significant impact on results concerning the effect of employee attitudes on performance. Models 1-3 utilize an archival measure of performance, improper

payment errors, while Model 4 utilizes a perceptual measure, based on asking respondents how they would rate the quality of work done by the work group. The drastically different findings are troubling. Examining performance as an archival measure suggests that while previous performance is very significant, essentially no attitudinal variables play a significant role in predicting performance. Still, the models are somewhat underspecified. It will be important to encourage further research on these issues with models including variables for program design and implementation features. When examining performance as a perceptual construct that references the workgroup, half of the attitudinal variables significantly predict performance. While the author's conceptualization of performance may underweight the significance of employee attitudes, it is also likely that the use of a perceptual measure for performance may *overweight* them. The likely possibility that the findings from Model 4 are false positives resulting from common method bias – the measurement errors for the dependent and attitudinal independent variables are likely to be highly correlated – should be a major concern for researchers examining variation in government programs' performance.

The second main finding is that previous performance dominates the variation in explaining performance, for the archival measure of payment errors. This finding provides suggestive evidence of the importance of *program design* and *locus of implementation*, relative to that of office management and bureaucratic values, when comparing performance *across* programs. Management researchers studying the effect of attitudes on performance should consider the ways in which design and implementation structure can build errors into a program. Future research could reveal that while increasing public employee motivation can lead to marginal improvement in program performance, carefully designed programs and implementation structures will create a stronger base, with the potential for even higher

performance; however, these suggestive findings will need to be rigorously examined in future studies as more data are collected and available.

Table 3.1: Theory of Bureaucratic Error: Error Categorization

	Purposeful	**Accidental**
Favorable	Favorable, purposeful error	Favorable, accidental error
Punitive	Punitive, purposeful error	Punitive, accidental error

Table 3.2: Summary Statistics 2004-2012	N	Mean	SD	Min	Max
Payment Error Rate	446	4.43	6.85	0.00	61.60
Amount Improper (Millions of Dollars)	446	1289.35	4056.50	0.00	30793
Amount improper in log form	413	4.18	2.75	-2.66	10.34
Outlays	446	28029.01	86509.76	0.20	695000.00
Within Organization Opportunities	396	2.98	0.12	2.60	3.25
Job Satisfaction	396	3.68	0.10	3.40	3.88
Pay Satisfaction	396	3.55	0.13	3.28	3.88
New ways of doing things encouraged	396	3.45	0.14	3.04	3.81
I like the kind of work I do	396	4.15	0.08	3.93	4.31
I recommend my organization as a good place to work	396	3.62	0.18	2.99	4.05
Supervisor support for life/work balance	299	3.92	0.19	3.29	4.25
I have sufficient resources	396	3.10	0.16	2.51	3.46
My workload is reasonable	396	3.33	0.10	2.89	3.66
Physical conditions	396	3.65	0.12	3.29	4.12
Promotions in my work unit are based on merit	396	2.88	0.17	2.47	3.25
My performance appraisal is a fair reflection of my performance	396	3.56	0.1	3.28	3.91
Supervisor discussions about my performance are worthwhile	396	3.47	0.11	3.22	3.86
I am held accountable for achieving results	396	4.00	0.08	3.71	4.23
Managers evaluate progress toward goals	396	3.54	0.15	3.21	4.00
Employees protected from health & safety hazards	396	3.80	0.17	3.24	4.18
Supervisors support employee development	396	3.59	0.11	3.30	3.91
Employees share job knowledge	396	3.78	0.09	3.60	4.04

Table 3.3: Results	(1) Error Rate	(2) $$$ Improper	(3) $$$ Improper Log	(4) Perception
Previous Year Performance	0.893**	0.954**	0.805**	0.447**
Outlays	-0.000	0.001	0.000+	-0.000*
Within Organization Opportunities	6.997	1396.420	-0.050	-0.127*
Job Satisfaction	21.325	252.602	-0.580	0.080
Pay Satisfaction	-3.930+	-167.257	-0.541	0.122**
New ways of doing things encouraged	-6.668	16.408	0.246	0.165**
I like the kind of work I do	-9.482+	-2378.755	-0.283	0.063
I recommend my organization as a good place to work	-12.104	168.901	-1.048	-0.009
Supervisor support for life/work balance	1.479	858.629+	0.994	-0.010
I have sufficient resources	2.746	-184.363	1.020	0.071*
My workload is reasonable	-0.924	-192.886	-0.822	-0.045
Physical conditions	9.629	-1282.783	0.746	0.013
Promotions in my work unit are based on merit	-3.071	-1394.074	-2.509	0.119**
My performance appraisal is a fair reflection of my performance	5.520	302.927	1.644	-0.058
Supervisor discussions about my performance are worthwhile	-13.522	-62.768	0.246	0.034
I am held accountable for achieving results	-5.250	2712.747	0.450	0.198**
Managers evaluate progress toward goals	4.690	18.405	1.023	-0.100**
Employees protected from health & safety hazards	-1.272	657.173	0.568	-0.035
Supervisors support employee development	7.934	-587.971	-0.565	-0.126+
Employees share job knowledge	-0.575	637.652	0.439	0.088+
Observations	237	237	215	201
R-squared	0.6128	0.9596	0.8227	0.8795

Note: + p<0.10 * p<0.05 ** p<0.01

Table 3.4: Best & Worst Performers (2004-2011 average)

Best Performers		**Worst Performers**	
(504) Certified Development Company Debentures (CDC)	0%	Federal Protective Service	10.09%
7(a) Guaranty Approvals	0%	Fee Program	10.10%
CDC Loans Guarantied	0%	Unemployment Insurance (UI)	10.64%
USCG Contract Payments	0%	Child Care and Development Fund (CCDF)	11.98%
CBP Border Security Fencing	0.01%	Medicare Advantage (Part C)	12.79%
Insurances	0.02%	Universal Service Fund - High Cost	13.32%
Aviation Security - Payroll	0.03%	Supplies and Materials	13.60%
Research and Education Grants	0.04%	State Home Per Diem Grants	13.68%
Federal Employee Life Insurance Program	0.06%	Non-insured Assistance	13.70%
Federal Employees Compensation Act (FECA)	0.10%	International Information Program	14.69%
Community Development Block Grant	0.12%	Children's Health Insurance Program (CHIP)	14.70%
Workforce Investment Act (WIA)	0.18%	School Lunch	16.31%
FTA Formula Grants Program	0.22%	Disaster Assistance Loans	18.31%
Clean & Drinking Water State Revolving Fund (SRF)	0.22%	School Breakfast	24.88%
USAID Twenty Seven Program Areas	0.22%	Earned Income Tax Credit (EITC)	25.19%
CBP Custodial Refund & Drawback	0.27%		
USAID Cash Transfers	0.30%		
Federal Employee Health Benefit Program	0.34%		
FAA Airport Improvement Program	0.35%		
Title I - Grants to States	0.35%		
Federal Employee Retirement Programs	0.36%		
USCG Active Duty Military Payroll	0.40%		
Transit Security Grants Program	0.41%		
Loan Guaranty	0.46%		
Farm Security & Rural Investment	0.49%		
Old Age, Survivors & Disability Insurance (OASDI)	0.51%		
USAID Grants, Contracts and Cooperative Agreements	0.55%		

REFERENCES

Adams, G. B., & Bullock, J. B. (2013). Bureaucratic Error: The Case of the United States Unemployment Insurance Program. *European Group on Public Administration Conference*.

Andrews, R., & Brewer, G. A. (2013). Social Capital, Management Capacity and Public Service Performance. *Public Management Review*, *15*(1), 19–42.

Avellaneda, C. N. (2012). Mayoral Decision-Making: Issue Salience, Decision Context, and Choice Constraint? An Experimental Study with 120 Latin American Mayors. *Journal of Public Administration Research and Theory*.

Baron, J. N., & Bielby, W. T. (1980). Bringing the firms back in: Stratification, segmentation, and the organization of work. *American Sociological Review*, *45*(5), 737–765.

Baron, J. N., & Bielby, W. T. (1984). The organization of work in a segmented economy. *American Sociological Review*, 454–473.

Bazerman, M. H., & Moore, D. A. (2008). Judgment in managerial decision making.

Beaton, B. J. (2008). Walking the Federalist Tightrope: A National Policy of State Experimentation for Health Information Technology. *Columbia Law Review*, 1670–1717.

Bellman, R. E., & Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, *17*(4), B–141.

Bendor, J. B. (1985). *Parallel systems: Redundancy in government*. University of California Pr.

Bertelli, A. M., Mason, D. P., Connolly, J. M., & Gastwirth, D. A. (2013). Measuring Agency Attributes with Attitudes Across Time: A Method and Examples Using Large-Scale Federal Surveys. *Journal of Public Administration Research and Theory*,

Blau, P. M. (1969). *The dynamics of bureaucracy: A study of interpersonal relations in two government agencies*. University of Chicago Press.

Boyne, George A., Oliver James, Peter John, and Nicolai Petrovsky. "Top Management Turnover and Organizational Performance: A Test of a Contingency Model." *Public Administration Review* 71, no. 4 (2011): 572–581.

Brewer, G. A. (2005). In the Eye of the Storm: Frontline Supervisors and Federal Agency Performance. *Journal of Public Administration Research and Theory*, *15*(4), 505–527.

Bullock, J. B. (2014a). A Theory of Bureaucratic Error. *Working Paper*.

Bullock, J. B. (2014b). Performance in the Public Sector: A Look at Federal Improper Payments. *Working Paper*.

Bullock, J. B., Stritch, J. M., & Rainey, H. G. (2014). Do sector differences hold across countries? An international comparison of public and private employees' motives and attitudes. *Working Paper*.

Bullock, J. B., Wenger, J. B., & Wilkins, V. M. (2014). Attitudes About Hard Work: A Global Perspective on the Beliefs of Government Employees. *International Public Management Journal*, *17*(1), 1–20.

Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, *12*(1), 9–30.

Clark, P. B., & Wilson, J. Q. (1961). Incentive systems: A theory of organizations. *Administrative Science Quarterly*, 129–166.

Chisholm, D. (1989). *Coordination without hierarchy: Informal structures in multiorganizational systems*. University of California Pr.

Carlson, D. E., Cowen, J. M., & Fleming, D. J. (2013). Third-Party Governance and Performance Measurement: A Case Study of Publicly Funded Private School Vouchers. *Journal of Public Administration Research and Theory*

Cho, Yoon Jik, and Evan J. Ringquist. "Managerial Trustworthiness and Organizational Outcomes." *Journal of Public Administration Research and Theory* 21, no. 1 (2011): 53–86.

Christensen, R. K., Szmer, J., & Stritch, J. M. (2012). Race and Gender Bias in Three Administrative Contexts: Impact on Work Assignments in State Supreme Courts. *Journal of Public Administration Research and Theory*.

Chun, Y. H., & Rainey, H. G. (2005a). Goal ambiguity and organizational performance in US federal agencies. *Journal of Public Administration Research and Theory*, *15*(4), 529–557.

Chun, Y. H., & Rainey, H. G. (2005b). Goal ambiguity in US federal agencies. *Journal of Public Administration Research and Theory*, *15*(1), 1–30.

Clark, P. B., & Wilson, J. Q. (1961). Incentive systems: A theory of organizations. *Administrative Science Quarterly*, 129–166.

Coggburn, J. D., & Schneider, S. K. (2003). The quality of management and government performance: An empirical analysis of the American states. *Public Administration Review*, *63*(2), 206–213.

Cyert, R. M., & March, J. (1963). A behavioral theory of the firm. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

Department of Labor. (2013, April). Unemployment Compensation Federal-State Partnership. Department of Labor.

Downs, A. (1964). *Inside bureaucracy*. Little, Brown Boston.

Dutton, S., De Pinto, J., Salvanto, A., & Backus, F. (2013, November 20). Poll: Obamacare support, Obama approval sink to new lows. *CBSNEWS*. Retrieved from http://www.cbsnews.com/news/poll-obamacare-support-obama-approval-sink-to-new-lows/

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

Favero, N., & Bullock, J. B. (2014). Identifying, Evaluating, and Remdying the Common Source Bias Problem in Public Administration Research. *Working Paper*.

Fernandez, S., Cho, Y. J., & Perry, J. L. (2010). Exploring the link between integrated leadership and public sector performance. *The Leadership Quarterly*, *21*(2), 308–323.

Fernandez, S., & Moldogaziev, T. (2013). Using Employee Empowerment to Encourage Innovative Behavior in the Public Sector. *Journal of Public Administration Research and Theory*, *23*(1), 155–187.

Frenkel, S. (1999). *On the front line: Organization of work in the information economy*. Cornell University Press.

Gallo, N., & Lewis, D. E. (2012). The Consequences of Presidential Patronage for Federal Agency Performance. *Journal of Public Administration Research and Theory*, *22*(2), 219–243.

Gardner, J. A. (1995). States-as-Laboratories Metaphor in State Constitutional Law, The. *Val. UL Rev.*, *30*, 475.

Gilmour, John B., and David E. Lewis. “Does Performance Budgeting Work? An Examination of the Office of Management and Budget’s PART Scores.” *Public Administration Review* 66, no. 5 (2006): 742–752.

Gilmour, J. B., & Lewis, D. E. (2006). Political appointees and the competence of federal program management. *American Politics Research, 34*(1), 22–50.

Gormley, W. T. (1996). Regulatory Privatization: A Case Study. *Journal of Public Administration Research and Theory*, *6*(2), 243–260.

Greenberg, D. H., Linksz, D., & Mandell, M. (2003). *Social experimentation and public policymaking*. The Urban Insitute.

Groves, T. (1973). Incentives in teams. *Econometrica: Journal of the Econometric Society*, 617–631.

Handler, J. F. (1992). Discretion: Power, quiescence, and trust. *The Uses of Discretion*, 331–360.

Heinrich, C. J. (2012). How credible is the evidence, and does it matter? An analysis of the program assessment rating tool. *Public Administration Review*, *72*(1), 123–134.

Hvidman, U., & Andersen, S. C. (2014). Impact of Performance Management in Public and Private Organizations. *Journal of Public Administration Research and Theory*, *24*(1), 35–58.

Huselid, M. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, *38*(3), 635–672.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* Free Press.

Johansen, M., & Zhu, L. (2014). Market Competition, Political Constraint, and Managerial Practice in Public, Nonprofit, and Private American Hospitals. *Journal of Public Administration Research and Theory*, *24*(1).

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review*, *16*(2), 366–395.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*(3), 376.

Jung, C. S. (2011). Organizational goal ambiguity and performance: Conceptualization, measurement, and relationships. *International Public Management Journal*, *14*(2), 193–217.

Jung, C. S. (2014). Why Are Goals Important in the Public Sector? Exploring the Benefits of Goal Clarity for Reducing Turnover Intention. *Journal of Public Administration Research and Theory*, *24*(1), 209–234.

Jung, C. S., & Rainey, H. G. (2009). Program types, goal ambiguity, and performance in US federal programs. In *Public Management Research Association bi-annual conference, Columbus, OH, October* (pp. 1–3).

Jung, C. S., & Rainey, H. G. (2011). Organizational goal characteristics and public duty motivation in US federal agencies. *Review of Public Personnel Administration*, *31*(1), 28–47.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–291.

Keiser, L. R. (1999). State bureaucratic discretion and the administration of social welfare programs: The case of social security disability. *Journal of Public Administration Research and Theory*, *9*(1), 87–106.

Kim, S. (2005). Individual-level factors and organizational performance in government organizations. *Journal of Public Administration Research and Theory*, *15*(2), 245.

Klein, G. A., Orasanu, J. E., Calderwood, R. E., & Zsambok, C. E. (1993). Decision making in action: Models and methods. In *This book is an outcome of a workshop held in Dayton, OH, Sep 25–27, 1989.* Ablex Publishing.

Laffont, J.-J., & Tirole, J. (1991). The politics of government decision-making: A theory of regulatory capture. *The Quarterly Journal of Economics*, *106*(4), 1089–1127.

Landau, M. (1969). Redundancy, rationality, and the problem of duplication and overlap. *Public Administration Review*, *29*(4), 346–358.

Lee, S.-Y., & Whitford, A. B. (2008). Exit, Voice, Loyalty, and Pay: Evidence from the Public Workforce. *Journal of Public Administration Research and Theory*, *18*(4), 647–671.

Lens, V. (2009). Seeking Justice: Citizens' Use of Fair Hearings to Correct Errors in Public Welfare Bureaucracies. *Journal of Public Administration Research and Theory*, *19*(4), 817–837.

Lewis, D. E. (2008). *The politics of presidential appointments: Political control and bureaucratic performance*. Princeton University Press.

Linder, S. H., & Peters, B. G. (1984). From social theory to policy design. *Journal of Public Policy*, *4*(03), 237–259.

Linder, S. H., & Peters, B. G. (1987). A design perspective on policy implementation: The fallacies of misplaced prescription. *Review of Policy Research*, *6*(3), 459–475.

Lipsky, M. (1980). *Street-level bureaucracy*. Russell Sage Foundation.

Lynn, L. E., Heinrich, C. J., & Hill, C. J. (2000). Studying governance and public management: Challenges and prospects. *Journal of Public Administration Research and Theory, 10*(2), 233–262.

May, P. J. (1992). Policy learning and failure. *Journal of Public Policy*, 331–354.

Maynard-Moody, S. W., & Musheno, M. C. (2003). *Cops, teachers, counselors: Stories from the front lines of public service*. University of Michigan Press.

Maynard-Moody, S., & Musheno, M. (2003). *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service*. University of Michigan Press.

Maynard-Moody, S., & Musheno, M. (2012). Social Equities and Inequities in Practice: Street-Level Workers as Agents and Pragmatists. *Public Administration Review*, *72*(s1), S16–S23.

Meier, K. J., & O'Toole, L. J. (2001). "Managerial Strategies and Behavior in Networks: A Model with Evidence from US Public Education,." *Journal of Public Administration Research and Theory*, *11*(3), 271–93.

Meier, K. J., & O'Toole, L. J. (2002). Public management and organizational performance: The effect of managerial quality. *Journal of Policy Analysis and Management*, *21*(4), 629–643.

Meier, K. J., & O'Toole, L. J. (2002). Public management and organizational performance: The effect of managerial quality. *Journal of Policy Analysis and Management*, *21*(4), 629–643.

Meier, K. J., & O'Toole, L. J. (2003). Public management and educational performance: The impact of managerial networking. *Public Administration Review*, *63*(6), 689–699.

Meier, K. J., & O'Toole, L. J. (2005). "Managerial Networking: Issues of Measurement and Research Design." *Administration & Society, 37*(5), 523–541.

Meier, K. J., & O'Toole Jr, L. J. (2013). I Think (I Am Doing Well), Therefore I Am: Assessing the Validity of Administrators' Self-Assessments of Performance. *International Public Management Journal*, *16*(1), 1–27.

Meier, K. J., & O'Toole, L. J. (2013). Subjective organizational performance and measurement error: Common source bias and spurious relationships. *Journal of Public Administration Research and Theory*, *23*(2), 429–456.

Meier, K. J., O'Toole, L. J., Boyne, G. A., & Walker, R. M. (2007). Strategic Management and the Performance of Public Organizations: Testing Venerable Ideas against Recent Theories. *Journal of Public Administration Research and Theory*, *17*(3), 357–377.

Miller, G. (2000). Above Politics: Credible Commitment and Efficiency in the Design of Public Agencies. *Journal of Public Administration Research and Theory*, *10*(2), 289–328.

Miller, G. J., & Whitford, A. B. (2007). The Principal's Moral Hazard: Constraints on the Use of Incentives in Hierarchy. *Journal of Public Administration Research and Theory, 17*(2), 213 –233.

Moynihan, D. P., & Lavertu, S. (2012). Cognitive Biases in Governing: Technology Preferences in Election Administration. *Public Administration Review*, *72*(1), 68–77.

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., & Pandey, S. K. (2005). Testing How Management Matters in an Era of Government by Performance Management. *Journal of Public Administration Research and Theory*, *15*(3), 421–439.

Mosher, F. C. (1968). *Democracy and the public service* (Vol. 53). New York.

Moynihan, D. P., & Lavertu, S. (2012). Cognitive Biases in Governing: Technology Preferences in Election Administration. *Public Administration Review*, *72*(1), 68–77.

Niskanen, W. A. (1971). *Bureaucracy and representative government*. Transaction Books.

Nutt, P. C. (2006). Comparing Public and Private Sector Decision-Making Practices. *Journal of Public Administration Research and Theory*, *16*(2), 289 –318.

O'Reilly, C. A. (1983). The use of information in organizational decision making: A model and some propositions. *Research in Organizational Behavior*.

O'Toole, L. J. (2000). Research on policy implementation: Assessment and prospects. *Journal of Public Administration Research and Theory*, *10*(2), 263–288.

O'Toole, L. J. (2000). Research on policy implementation: Assessment and prospects. *Journal of Public Administration Research and Theory*, *10*(2), 263–288.

O'Toole, L. J., & Meier, K. J. (1999). Modeling the impact of public management: Implications of structural context. *Journal of Public Administration Research and Theory*, *9*(4), 505.

O'Toole, L. J., & Meier, K. J. (2003). Plus ça change: Public management, personnel stability, and organizational performance. *Journal of Public Administration Research and Theory*, *13*(1), 43–64.

O'Toole, L. J., & Meier, K. J. (2004). Public management in intergovernmental networks: Matching structural networks and managerial networking. *Journal of Public Administration Research and Theory*, *14*(4), 469.

Pauly, M. V. (1974). Overinsurance and public provision of insurance: the roles of moral hazard and adverse selection. *The Quarterly Journal of Economics*, *88*(1), 44–62.

Perry, J. L., & Rainey, H. G. (1988). The public-private distinction in organization theory: A critique and research strategy. *Academy of Management Review*, *13*(2), 182–201.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569.

Plous, S. (1993). *The psychology of judgment and decision making.* Mcgraw-Hill Book Company.

Provan, Keith G., and H. Brinton Milward. "A Preliminary Theory of Interorganizational Network Effectiveness: A Comparative Study of Four Community Mental Health Systems." *Administrative Science Quarterly* 40, no. 1 (1995).

Rabe, B. (2007). Environmental policy and the Bush era: The collision between the administrative presidency and state experimentation. *Publius: The Journal of Federalism*, *37*(3), 413–431.

Radin, B. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Georgetown University Press.

Rainey, H. G. (1983). Public agencies and private firms. *Administration & Society*, *15*(2), 207.

Rainey, H. G. (2009). *Understanding and managing public organizations*. Jossey-Bass.

Rainey, H. G., & Bozeman, B. (2000). Comparing public and private organizations: Empirical research and the power of the a priori. *Journal of Public Administration Research and Theory*, *10*(2), 447.

Riccucci, N. M., & Meyers, M. K. (2004). Linking passive and active representation: The case of frontline workers in welfare agencies. *Journal of Public Administration Research and Theory*, *14*(4), 585–597.

Ryu, S., Wenger, J. B., & Wilkins, V. M. (2012). When Claimant Characteristics and Prior Performance Predict Bureaucratic Error. *The American Review of Public Administration*, *42*(6), 695–714.

Sabharwal, M. (2013). From Glass Ceiling to Glass Cliff: Women in Senior Executive Service. *Journal of Public Administration Research and Theory*.

Sandfort, J., Selden, S. C., & Sowa, J. E. (2008). Do government tools influence organizational performance? Examining their implementation in early childhood education. *The American Review of Public Administration*, *38*(4), 412–438.

Schneider, A., & Ingram, H. (1993). Social construction of target populations: Implications for politics and policy. *The American Political Science Review*, *87*(2), 334–347.

Schneider, M. (2004). Performance Management by Culture in the National Labor Relations Board's Division of Judges and the German Labor Courts of Appeal. *Journal of Public Administration Research and Theory*, *14*(1), 19–32.

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

Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review*, *69*(4), 493–513.

Sosin, M. R. (2012). Decentralization, Devolution, Financial Shortfalls, and State Priorities in Service Programs in the Early 2000s. *Journal of Public Administration Research and Theory*, *22*(4), 701–730.

Soss, J. (1999). Lessons of welfare: Policy design, political learning, and political action. *American Political Science Review*, 363–380.

Soss, J., Fording, R., & Schram, S. F. (2011). The organization of discipline: From performance management to perversity and punishment. *Journal of Public Administration Research and Theory*, *21*(suppl 2), i203–i232.

Thompson, F. J. (1976). Minority Groups in Public Bureaucracies Are Passive and Active Representation Linked? *Administration & Society*, *8*(2), 201–226.

Tosi, H. L., Katz, J. P., & Gomez-Mejia, L. R. (1997). Disaggregating the agency contract: The effects of monitoring, incentive alignment, and term in office on agent decision making. *Academy of Management Journal*, *40*(3), 584–602.

Trevino, L. K. (1986). Ethical decision making in organizations: A person-situation interactionist model. *Academy of Management Review*, *11*(3), 601–617.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

U.S. Department of Labor. (2013, April). Unemployment Compensation: Federal-State Partnership. US Dept. of Labor, Employment and Training Administration, Office of Unemployment Insurance.

Vandenabeele, W. (2009). The mediating effect of job satisfaction and organizational commitment on self-reported performance: more robust evidence of the PSM—performance relationship. *International Review of Administrative Sciences, 75*(1), 11–34.

Walker, Richard M., and Rhys Andrews. “Local Government Management and Performance: A Review of Evidence.” *Journal of Public Administration Research and Theory* (2013).

Wart, M. V. (2003). Public-Sector leadership theory: An assessment. *Public Administration Review*, *63*(2), 214–228.

Wenger, J. B., & Wilkins, V. M. (2009). At the discretion of rogue agents: How automation improves women’s outcomes in unemployment insurance. *Journal of Public Administration Research and Theory*, *19*(2), 313–333.

Wilkins, V. M., & Keiser, L. R. (2006). Linking passive and active representation by gender: The case of child support agencies. *Journal of Public Administration Research and Theory*, *16*(1), 87–102.

Wilson, W. (1887). The study of administration. *Political Science Quarterly*, *2*(2), 197–222.

Yang, K., & Kassekert, A. (2010). Linking Management Reform with Employee Job Satisfaction: Evidence from Federal Agencies. *Journal of Public Administration Research and Theory*, *20*(2), 413–436. doi:10.1093/jopart/mup010

Zajac, G. (1996). Beyond Hammurabi: A Public Service Definition of Ethics Failure. *Journal of Public Administration Research and Theory*, *6*(1), 145–160.

Zeleny, M., & Cochrane, J. L. (1982). *Multiple criteria decision making* (Vol. 25). McGraw-Hill New York.