Dialogue Theory: The Language of Performance Regulation

Donald P. Moynihan
Assistant Professor of Public Affairs
La Follette School of Public Affairs
University of Wisconsin-Madison
Email: dmoynihan@lafollette.wisc.edu

This paper is an excerpt from *Rethinking Performance Management,* a book to be published with Georgetown University Press. Please do not cite without permission. Questions and comments welcome.

Abstract

Dialogue Theory: The Language of Performance Regulation

Performance information has become the language for how public services are regulated. This paper examines the limitations of such an approach, by presenting a theory of performance information-use I refer to as dialogue theory. Dialogue theory assumes that performance information is subjective, ambiguous, is shaped by those who present it, is interpreted via individual and institutional biases, and is used as a form of persuasion. In short, performance is constructed by interested actors who seek to persuade others by using information that supports their arguments. To illustrate the validity of dialogue theory I examine recent efforts by the Bush Administration to tie program performance assessments to the budget. While the White House sees such reforms as transparent and systematic, it faces criticism that it is subjective, reflecting the perspective of central agencies rather than those of agencies and Congress. Discussion of the use of such evaluation tools also illustrates their ambiguity, as actors in different interpret performance information in different ways.

Introduction

A central argument of the New Public Management was that performance information could and should be used to inform policy and management decisions. Given the current emphasis on evidence-based policy, and new program assessment tools, one might be tempted to believe that performance information is indeed having an impact. This paper examines the limitations of regulating performance by relying performance information as the language of regulation. I refer to this perspective as dialogue theory.

Dialogue theory emphasizes the ambiguity of performance information and the loose link between such information and decisions. Dialogue theory assumes that performance information is subjective, ambiguous, is shaped by those who present it, is interpreted via individual and institutional biases, and is used as a form of persuasion. In short, performance information is not a natural phenomenon; rather, it is constructed by interested actors who seek to persuade others by using information that supports their arguments.

Dialogue theory suggests that performance information does not necessarily result in clearer decisions if the actors involved cannot agree on what it tells them about current performance, budgets decisions or management. Different actors may disagree on how to determine performance, and what performance information means for resource allocation. For instance, a poorly performing program might face arguments that it needs to be cut, receive more funding, or be reorganized. Each option applies a logical interpretation of the same piece of data, underlining the ambiguity of interpreting performance.

Political and institutional roles will shape how political actors interpret, present and use performance information. Those with power, legitimacy and resources have a strong opportunity to shape how discourse is constructed in ways that sustain their authority and promotes their perspective.

I use the experience of the Program Assessment Rating Tool (PART) to illustrate dialogue theory. PART is an initiative of the U.S. Office of Management and Budget to rate the performance of all federal programs over a five year period. The paper draws on interviews with federal government budgeters who have worked on PART, and also presents an initial experiment using graduate students to test the basic assumptions of dialogue theory. The main goal of this paper is to extend a nascent theory for examining how performance information is used by public decision-makers to regulate performance. While the PART initiative is particular to the US, the basic issues of central-agency tensions in defining performance are common in any effort by the center to regulate agency performance.

Dialogue Theory

Proposals for the integration of performance data and budgeting tend to assume that performance information is objective, standardized, indicative of actual performance, consistently understood, and prompts a consensus about how a program is performing and how it should be funded. My experience in examining the use of performance information among U.S. state governments suggested the need for a contrary set of assumptions, which form the basis of dialogue theory:

- Performance information is rarely comprehensive.
- Performance information is ambiguous.
- Performance information is subjective.
- Production of performance information does not guarantee use.

- Institutional affiliation and individual beliefs will affect selection, perception and presentation of performance information.
- The context of dialogue will affect the ability to use performance information to develop solutions.

Dialogue theory assumes that simply because performance information exists, there is no guarantee that it is used. Whether it is used, and how it is used, depends on the motivations of potential users, and their ability to use performance information to further their goals. Because performance information will sometimes help to advance an argument, it is likely to be used, but not in the simplistic or determinative way assumed by the strict theory of performance budgeting. Dialogue theory points to the ambiguity inherent in interpreting performance information. There is likely to be no single definitive approach to a) interpreting what performance information means and b) how performance information directs decisions. The meaning of performance information is constructed, and therefore the same performance information can support different arguments.

Previous theories provide support for the assumptions outlined above. First, the study of organizations and decisionmaking provides evidence of the ambiguity inherent in organizational life, the contested nature of discourse, and the potential for rival interpretations of performance information. Second, we know from the study of the policy process that the design of the budget process creates incentives for particular actors to advance arguments that reflect their institutional role and context, enhancing the potential for disagreement.

Ambiguity in Organizational Life

The study of organizational life reveals its deep-seated ambiguity.[i] Ambiguity is "that state of having many ways of thinking about the same circumstances or phenomena."[ii] Ambiguity is likely to occur in issues where objectives or issue-definition is unclear, where there is a lack of clarity on causal mechanisms between organizational actions and outcomes, where it is difficult to interpret the past, and where the pattern of individual participation in different decisions is uncertain and changing.[iii]

While more information might reduce uncertainty, it will not eliminate ambiguity, since ambiguity is created by different perspectives rather than a lack of information. Feldman suggests that ambiguous issues must be interpreted, a process by which actors give meaning to the issue: "Resolution is a matter of agreement rather than proof. To the extent that resolution occurs, it comes from shared understandings, not factual information."[iv] While information helps actors interpret a policy issue, it does not necessarily foster consensus on decisions. According to March:

> Organizational information processing seems to be driven less by uncertainty about the consequences of specific decision alternatives than by lack of clarity about how we talk about the world – what alternatives there are and how they are related to stories we think we understand, how we describe history and how we interpret it.[v]

Another approach to studying organizational life is to study its discourse, or how texts are created and exchanged between organizational actors. The literature on organizational discourse also focuses the ambiguity of texts, and assumes that the attitudes and behaviors of organizational actors are shaped by the discursive practices in the organization. Meaning is socially constructed. Each text is rife with unspoken assumptions, although each text is also susceptible to rival interpretations.[vi]

Differences in perspective, or subjectivity, is a natural result of ambiguity. Deciphering what constitutes performance can be difficult. Such variation in interpretation will increase when we face even trickier questions. What does performance mean? What are the next steps for the program? How should we budget on the basis of performance information? While performance information tells us something about a program, the data itself does not answer these questions.[vii] Performance data, or simplified assessments of performance data, fail to tell us:

- why performance did or did not occur
- the context of performance
- how implementation occurred
- an understanding of outside influences on performance
- how to choose which program measure is a priority

The absence of this information makes it difficult to determine what performance actually means. Even if you and I can agree on an appropriate performance measure, and can also agree on what the data means, we may still disagree on what to do next. A variety of plausible and logical different possible decision options may exist.[viii] One person might decide that the type of action appropriate for the program is related to funding, where another sees a management problem. Even if two individuals agree that the performance information should influence resource allocation, performance data does not tell us how. The classic example is a program with poor performance. Should funding be cut and the program abandoned as a failure, or should more resources be provided to help make the program a success? The ability to answer this question depends a great deal on our understanding of why performance failed to occurred and whether it can be remedied, issues that are also subject to disagreement. But our decision is also dependent on values. We are less likely to abandon programs that have an important purpose.

Performance information also cannot help us trade-off between multiple program and agency goals. Even if we accurately understood in advance the cost and outcomes of programs, that still does not provide a common basis for comparison since our willingness to fund services and specific levels of performance will depend on values. Indeed, the introduction of performance information simply adds related contextual questions: how do we know if more money will improve performance or be wasted? More broadly, how do we understand performance information, and how do we relate it to action?

The Role of Roles

Intelligent individuals with no particular axe to grind can disagree with one another about what performance information is appropriate, what it means, and what action should be taken as a result. The potential for ambiguity to lead to disagreement becomes much stronger in a political setting. In politics, actors play roles. They are representatives of ideologies, parties, programs, groups of citizens, organizations and political institutions. This will make it more likely that the ambiguity of performance information will be exposed.

How an individual interprets and understand information will be shaped by their role. March and Olsen argue that individual behavior and preferences are endogenous to the organizational context of the individual.[ix] Conditions of ambiguity and disagreement will therefore be exacerbated in political institutions that are designed to check one another and represent opposing viewpoints. The political arena is characterized by battles between rival actors to assert hegemony over issue definition.[x] Information selection and use occurs in the context of different beliefs, preferences, and cognitive processes, and will reflect organizational power and politics. Information providers will try to shape outcomes by choosing what information will be collected and highlighted. Each measure is representative of values and accompanied by the assumption that the organization should be making efforts that will have an impact on the measure. Or, as March succinctly puts it: “[I]nformation in organizations is not innocent.”[xi]

Deborah Stone makes almost exactly the same point “No number is innocent, for it is impossible to count without making judgments about categorization.”[xii] Stone describes how interpretations of policies are constructed into narrative stories of how the world works. Policy problems are defined and communicated in symbolic terms, and the most important aspect of symbols is their ambiguity. Among the most frequently used symbols in establishing political narratives are numbers: "Numbers in politics are measures of human activities, made by human beings, and intended to influence human behavior. They are subject to conscious and unconscious manipulation by the people being measured, the people making the measurements, and the people who interpret and use measures made by others."[xiii] There can be political fights about numbers in terms of what is measured, how it is measured, what the measurement means. Programs often have multiple goals that sometimes conflict with one another. As a result, different measures of program performance can change how programs are considered, as suggested by Heinrich and Fournier: “The important question is not ‘What structure or type of program treatment works best?’ but ‘What aspects of organization and treatment services work for whom, in what context, and toward what treatment goals?’”[xiv]

Numbers can be used to tell stories, and sometimes to suggest that complex phenomena can be precisely defined and controlled. Numbers are powerful and offer the user a sense of authority because we associated data with objectivity and professionalism. The same number can have multiple meanings, which makes them contestable.[xv] To make performance information understandable, it needs advocates. Advocates direct attention to particular pieces of information, make sense of what it means in terms of resource needs, offering decision suggestions to their audiences. They construct narratives and stories around what the performance information means and what it suggests should be done.

Research in psychology also points to a confirmation bias in the selection and use of information in decisionmaking – once decisionmakers have made up their mind on an issue, they will seek out information that supports their views, and discount contrary information.[xvi] This insight provides helps us to understand why preferences formed by institutional norms become hard to displace, influencing the selection and perception of performance information. Confirmation bias can even affect the perception of the same information, with research finding that "two people with initially conflicting views can examine the same evidence and both find reasons for increasing the strength of their existing opinions."[xvii]

The organizational discourse literature offers some insights in how roles shape how texts are created and whether they become embedded in discourse, i.e. widely used by a target audience. Organizational actors are also more likely to create texts that reflect actions that burnish their image, and so it is not surprising that organizations are likely to produce performance data in a way that makes them look effective or supports some other type of organizational argument. However, the existence of competing texts and rival perspectives lowers the potential for acceptance. Therefore, in an environment where actors represent different views and offer different interpretations, there is less chance of any single interpretation dominating. This is a useful observation given the pluralistic tendencies of public dialogue.[xviii]

Political roles will shape how political actors interpret, present and use performance information. Conflict in interpretation is will also be fostered by the norms and incentives associated with particular roles in the budget process, and observers of the budget process have noted that way that individuals tend to conform to role expectations.[xix]. Incrementalist theory described these roles.[xx] Agency staff are likely to be advocates, pointing to performance data that makes their program look good, discounting less flattering data, and arguing that performance results calls for more resources. Oversight committees will mix advocacy with a desire to exercise control over the agency. Central agency staff are expected to manage and verify agency claims about performance and the implications for resources. One empirical example of how roles shape perspectives comes from a survey of state budgeters and agency practitioners. Asked what were significant problems in performance measurement, 40% of agency practitioners point to “too many factors affect the results trying to be achieved”, while only 14% of state budgeters considered this to be a significant problems.[xxi] Clearly agency staff are more concerned with considering how outside factors shape performance results than central agency staff.

The importance of roles suggests that the ability to foster problem-solving dialogue is reduced when multiple actors from different institutional settings are engaged, because such a diversity of views reduces the ability to foster consensus. All else equal, dialogue routines which are less institutionally diverse are more likely to use performance information to solve problems. For this reason, traditional budget routines have a limited ability to create dialogue which will generate solutions. Agency staff or stakeholders presenting information to the central budget office or testifying before the legislature have good reason to act defensively, and will rarely address operational questions. In such situations, the trademark of dialogue will be advocacy. Information will be used strategically by advocates who are fully aware that they argue in a decision environment does not closely resemble a strict performance budgeting process. Decisionmakers who receive budget information from agencies lack the time, interest and capacity to make decisions on resource allocation based on performance information. The budget cycle is too short, and budgeters have enough to do to make budget decisions incrementally, even with all of the heuristics that are associated with this approach. It is nearly impossible for them to make judgments on what mounds of performance information indicates for a particular functional area. This is true even if only a handful of performance measures are reported. There will never be enough information to substitute for the expertise and knowledge to enable substantive judgments to be made, and there will always be too much information for human cognitive processes to deal with.

Of course these advocates are biased in their assessments, but all parties of aware of this, and it does not mean they are wrong. In fact, the greater the degree to which they can use the performance information to support their position, the less biased and more rational they appear.

A theory should be able to offer some falsifiable hypotheses. Based on the assumptions of dialogue theory presented above, and the supporting literature, the following hypotheses about performance information and decisionmaking emerge:

- Different actors can examine the same programs and come up with competing, though reasonable, arguments for the performance of a program based on different data.
- Different actors can examine the same performance information and come up with competing, though reasonable, arguments for what the information means.

- Different actors can agree on the meaning of performance information/program performance and come up with competing solutions for what actions to take in terms of management and resources.
- Actors will select and interpret performance information consistent with institutional values and purposes.
- Forums where performance information is considered across institutional affiliations will see greater contesting of performance data.

The Program Assessment Rating Tool

The second half of this paper examines PART from the perspective of dialogue theory. The OMB formed the Performance Evaluation Team early in 2002 to help create a systematic program evaluation tool. The team identified a number of basic questions to ask of all programs and tried to identify what constituted acceptable evidence for answering these questions.[xxii] The team also set out the goal of evaluating all federal programs over a five year period.

PART is essentially a set of 30 standard questions that OMB budget examiners, in consultation with agency representatives, use to assess federal programs.[xxiii] It is not an original evaluation of programs, but an analysis of relevant evaluations that have been previously completed. Programs can be rated as ineffective, adequate, moderately effective or effective. If in OMB's judgment, programs fail to provide adequate information they may fall into a separate category, of results not demonstrated. Consistent with the goal of evaluating all federal programs over a five-year period, the first four years of PART has seen 793 programs (approximately 80 percent of the federal budget) have been "PARTed" with the following program ratings:

- 15 percent are Effective
- 29 percent are Moderately Effective
- 28 percent are Adequate
- 4 percent are Ineffective
- 24 percent are Results Not Demonstrated

The ratings are based on four weighted groupings of questions: program purpose and design (standard weight is 20 percent), strategic planning (10 percent), program management (20 percent) and program results (50 percent).[xxiv] Single-page summary assessments of each program accompanied the President's budget proposal to Congress.

The application of PART has also been relatively transparent. Detailed assessments are available on the OMB website for every program rated, describing the reasoning and information used by budget examiners to respond to each of the PART questions. The basic framework of the relationship between budget examiners and agencies is now open to the public in a way that it has never been before. Those that disagree with the evaluations can critique the rationale for each assessment. The OMB has worked hard to ensure consistency, designing standardized questions, training raters and providing a sixty page guide,[xxv] and even forming a team to conduct a consistency check on 10 percent of the assessments.

Evidence-Based Dialogue

OMB staff argue that PART is designed to elicit an evidence-based dialogue with agency staff, and to generate a variety of recommendations related to management, funding, program design and program assessment methods. The OMB sees PART as prompting evidence-based dialogue in a number of ways:

- *Third-party program review with a clear opinion*: PART seeks to deal with the perceived unreliability of agency self-reporting and the problem of information overload.
- *Greater emphasis on performance:* The basic focus of PART is performance, and the existence of the PART tool has increased discussion of program performance, at least in the budget preparation process within the executive branch.[xxvi]
- *The standard of proof for program performance can only be satisfied by positive evidence of results, rather than an absence of obvious failure:* If there is a lack of information about the success of a program, or if the OMB disagrees that the measures in place are valid, then a program will be rated as results not demonstrated. This raises the burden of proof to require actual evidence of program effectiveness.
- *The burden of proof for performance rests on agencies:* The OMB sees its role as assessing existing evidence on performance, but expects that agencies will collect and present this evidence.
- *Entire programs are evaluated on a regular basis:* PART is a major change to the traditional approach of budgeting, where programs are usually only given a thorough evaluation when first proposed. PART offers regular assessments of entire programs along with its funding recommendations, suggesting a willingness to make dramatic cuts where the evidence merits it.
- *The routine nature of PART creates an incentive to engage:* In addition to assessing all programs over a five-year period, reassessments of some programs take place when new evidence warrants it. However, even programs that are not fully reassessed are examined to see if they have followed up with management recommendations.

PART is used for OMB decisions. The most prominent impact of PART is its inclusion as part of the President's budget proposal. Within the OMB, the decisive debate that underpins this budget proposal is the Director's Review. Individual OMB budget examiners will discuss specific issues and offer recommendations, and in this context they will sometimes refer to PART assessments, particularly is there is a discussion about whether to cut or eliminate a program.

The Ambiguity of PART

One of the assumptions of dialogue theory is that performance information is ambiguous, holding different meanings for different individuals at the same time. OMB staff consider PART to be a systematic tool that, as much as possible, minimizes the subjectivity of the budget examination process. But raters will understand the same terms, apply standards and interpret data in different ways. In reviewing PART, performance measurement expert Harry Hatry noted: "Any ratings of program purpose and design, strategic planning, and program management are bound to be heavily subjective."[xxvii]

For instance, how does a rater evaluate the performance of a program that is categorized as results not demonstrated? The application of the results not demonstrated categorization is itself ambiguous. It is sometimes applied if there were inadequate performance information, but in other cases it is used if the OMB and agency simply failed to agree on what were appropriate long-term and annual performance measures for the program,[xxviii] or according to one budget examiner, if the program looked like it would receive a poor grade but the agency pleaded for more time to turn the program around. The understanding of what constituted "independent and quality evaluation" is also a source of disagreement between agencies and the OMB.[xxix]

Differences in interpretation of ambiguous terms will be greater between members with different institutional interests, but has also occurred among OMB staff. The wording of some important questions also gives rise to ambiguity, e.g. whether a program is “excessively” or “unnecessarily…redundant or duplicative of any other Federal, State, local, or private effort”; whether a program design is free of “major flaws”; whether a program’s performance measures “meaningfully” reflect the programs purpose; and whether a program has demonstrated “adequate” progress in achieving long-term performance goals. The use of such non-precise terms invites a variety of interpretations as to what they mean.[xxx] There has been inconsistency among OMB budget examiners on the meaning of terms such as “ambitious” “outputs”, “outcomes”, and “having made progress.”[xxxi] Agency officials have complained that OMB officials used different standards for defining what measures were outcome-oriented.[xxxii] Even efforts to standardize responses through the use of “yes/no” responses created problems, since different OMB staffers had different standards for what constituted a “yes.”[xxxiii]

The Subjectivity of PART

The basic ambiguity of information fosters subjectivity, and dialogue theory assumes that information is not objective, but selected, measured and presented in ways that reflect the biases and interests of the actors presenting it. According to the organizational discourse literature, a mutually constitutive relationship exists between power and discourse. Those with power, legitimacy and resources have a strong opportunity to shape how discourse is constructed in ways that sustain their authority and promotes their perspective.[xxxiv] The way in which information is presented will also affect acceptance, as genres of texts that are recognizable, interpretable and usable to others, or texts that draw on previously-established texts, are more likely to become embedded in discourse.[xxxv] Applying this insight to the public sector, central agency officials using the traditional budget format are in a relatively strong position to push performance information into the budget discourse. This OMB has done so with PART, although it remains to be seen the extent to which agency staff, and particularly Congress, will engage in OMB's version of evidence-based dialogue.

Performance management reforms are often described as politically neutral. "How can men account for so foolish a statement?" asks Wildavsky.[xxxvi] Such faith in the apolitical nature of performance management tools overlooks their centralizing tendencies. PART represents the interests, biases and views of the OMB. In tying PART to the budget process, OMB has used its most powerful tool to give PART prominence in key decisions. The OMB can define issues and shape interpretation by choosing what information are considered relevant and important, how to define a program's purpose, which programs are classified as succeeding or failing, and which programs are recommended for greater resources.

As Beryl Radin notes, many federal programs have multiple and conflicting goals, while the PART tool is designed on the assumption that there is only one dominant performance goal.[xxxvii] Clearly, having the power to decide what that goal will be while excluding other goals or measures will affect how success is understood. One OMB official noted that "there have been lots of disagreements between agency staff and OMB staff about what constituted an outcome." Agencies, and Congress, have even disagreed with the OMB on how they define programs. As Wildavsky noted many years ago, defining a program is a contested and important activity: "Programs are not made in heaven. There is nothing out there just waiting to be found. Programs are not natural to the world…There are as many ways of conceiving of programs as there are of organizing activities."[xxxviii]

If PART can help set the policy agenda by changing the nature of discussion about a program, then controlling PART represents a form of power and influence for the OMB and the White House. One OMB official suggested as much, saying: "I would like to believe that this is objective, but performance management is such a slippery slope, so its easy to take this stuff and make it political…PART is a gate-keeping process, it's a framing device. It affects how policies are defined and framed, which affects how issues are viewed. There is a lot of subjective bias even though it is an "objective" tool."

Not surprisingly, agencies sometimes disagree with the OMB about what performance information is relevant in evaluating programs. There are the traditional complaints about performance measurement: that it is inexact, that some programs are more difficult to measure than others, that there is often a significant lag-time between program treatment and outcome, and that performance oversight is more difficult for programs where service delivery is decentralized and the federal level exerts only partial control. Such criticisms are leveled toward every performance management reform, but that does not mean that they are not legitimate.

Agencies are also concerned about the OMB emphasis on outcome measures. Agencies tend to view their role as limited to what is under their control, and prefer output measures that reflect effort and inform process improvement efforts rather than ultimate outcomes that they many not fully, or even largely, control. [xxxix] The OMB sees agencies as responsible for national policies and therefore views outcome measures that are beyond the immediate control of the agency as more appropriate. A survey of federal managers found that 57 percent of non-Senior Executive Service (SES) managers and 61 percent of SES mangers felt they were being held accountable for outcomes, but significantly less - 38 percent and 40 percent, respectively – felt that they had the level of decision-authority to achieve these results.[xl] While PART has prompted greater attention to information that results from evaluation, the OMB and agencies have clashed on what sorts of evaluations are useful and credible, and frequently have different expectations about what sort of information evaluations should produce.[xli]

Agency staff are also likely to disagree with the OMB on the section of the PART that deals with the purpose of the program, since they are more likely than the OMB to view the purpose of their activities as self-evidently valuable. If agencies disagree with the decision of the budget examiner, they can appeal to political appointees in the OMB, but are generally not successful. While PART is a relatively transparent process, it is still dominated by the OMB, and largely excludes stakeholders and Congress. However, this does not prevent the agencies from using PART for advocacy purposes. Agencies have sometimes been successful in using PART to argue for greater resources, either for the program as a whole or for new evaluation efforts.[xlii]

In undertaking PART, OMB budget examiners consider themselves to be evaluators, and not just bean-counters. The scope of their influence on the agency enforces a norm that they need to consider management issues as well as budget issues.[xliii] PART offers another way to summarize and analyze evaluation information about programs, formalizing existing analytical practices that budget examiners already do. PART therefore reflects the professional norms of budget examiners and the OMB. To supporters of the OMB, those values include political neutrality and ensuring the probity and efficacy of public spending. To critics, the OMB and PART represents an effort to exert top-down control through one-size-fits-all tools, and a strong and increasing deference to Presidential preferences.

PART can also be seen as an OMB response to the more legislative-centered or agency level approach to performance management that characterized Government Performance and Results Act (GPRA) – legislation passed in 1993 that required agencies to develop strategic plans and measures. OMB raters often deem existing GPRA measures as unsatisfactory and have used PART recommendations to prioritize the type of performance information that the OMB values. In their analysis of the FY02-04 PART summaries, the Government Accountability Office (GAO) found that the majority of the recommendations the OMB have made with PART have been in the area of improving assessment methods – 52% of the 2002 recommendations, 59% in 2003 and 58% in 2004.[xliv] These recommendations will not directly improve the performance of these programs, but are designed to generate more or better performance data, making it easier for OMB examiners to assess whether a program satisfies the PART standard of success. The high percentage of assessment recommendations indicate OMB dissatisfaction with existing performance information used by agencies, as does the high number of programs that are classified as "results not demonstrated," but the GAO study found that these recommendations were the most frequent PART recommendation regardless of the deficiencies of the program, or the program's PART rating.[xlv]

Thus far, the attitude of Congress toward PART has ranged from indifference to occasional hostility.[xlvi] One OMB official said: "I don't want to say they don't care, but they don't care. They hold the purse strings and are proud of that role. They do not like to be told how to allocate money because of what a performance measure says." Congress already has performance information from GPRA, legislative analysts, and think tanks, as well as constituent views. One might expect that so much potential information would create a demand for a tool that promised to reduce the complexity of performance data, such as PART. Congress has resisted, partly due to messenger rather than the message. Legislators at both the state and federal level have shown a tendency to distrust the objectivity of any performance information reported by the executive branch. An OMB budget examiner conceded that "they may just disagree with the results.

Based on the same information, they come to different conclusions." Congressional staff have indicated that are suspicious of the PART tool, of how the OMB defines programs, what information they use to define performance, and the actual PART ratings.[xlvii]

Even if legislative committees accept PART analysis, they may believe that it is simply inconsistent with their priorities or the priorities of major stakeholders. After presenting PART information, another budget examiner recalled being told by an appropriations subcommittee clerk that "your proposals are for your priorities. We need to fund our priorities." One instance where committees may use PART recommendations is where interest groups are weak, and cuts are needed to provide money for other programs.

The OMB hopes that even if Congress disagrees with the evaluations, it will force a discussion between the executive and legislative branches in terms of performance. However, Congress is under no requirement to enter in such a discussion and thus far has largely refused to engage with the OMB. While the OMB has improved the supply of performance information, there is not yet, and perhaps will never be, a strong legislative demand for their product. At best, interest groups or agencies who find PART useful may bring it to the attention of legislative supporters, consistent with a fire-alarm approach to policymaking.[xlviii] It is worth noting that Republican control over both the Senate and the House provided the most favorable possible political context for PART to impact legislative decisionmaking. That it has had only marginal influence suggests the importance of institutional differences relative to partisan loyalties when it comes to public management reforms.

An Experimental Test of Dialogue Theory

Thus far this paper has reported on the actual experience of PART, providing evidence of the subjectivity that arises from cross-institutional conflicts about information. The remainder of the paper examines experimental evidence to assess the ambiguity of information, which dialogue theory hypothesizes will allow different actors to examine the same programs, and disagree on what performance data is relevant, and what the same performance data means. The experiment uses graduate students in public affairs programs at the Bush School of Government and Public Service at Texas A&M University, and at the La Follette School of Public Affairs at the University of Wisconsin-Madison. The experiment examines variation between PART program ratings and budget recommendations made by the OMB and students familiar with but not directly involved in the political process. The existence and reasons for such variation support the idea that different actors can employ logical warrants to interpret the same information differently.

In the experiment students were asked to take the position of a staff analyst for a Congressional committee and to analyze 3-5 programs overseen by that committee and evaluated by PART. Students were instructed to both examine the OMB's PART assessments but also to become an expert on the programs in their own right, undertaking research by examining the agency's strategic plan, performance report, website, agency, appropriate think tank reports, newspapers and professional magazines and other sources such as GAO reports. The students were asked whether they considered the OMB evaluation to be "fair and accurate, and for each program explain why you agree or disagree with the OMB solution," and to develop funding recommendations to provide to their committee.[xlix] Performing such an experiment is possible only because the PART process is so transparent, allowing students access to a high level of detail for each of the analytic categories that made up the PART score and rating, the process of reasoning behind the OMB responses, and the relative importance of these factors.

It should be noted that the purpose of the experiment is not to replicate the decision of an OMB or congressional analyst specifically, but to examine whether individuals with general policy analytical skills can look at the performance information for a specific program and come to different – though logical and defensible – assessments. Other research has relied on graduate students in public affairs as basis for generalizing to the behavior of public officials, particularly in the area of information processing and decisionmaking.[1] Such students have training in policy analysis and public management, largely intend to make their career in public policy, have worked on practical projects with real public organizations, and are required to intern with public organizations. Many frequently have previous experience working in government, and many will occupy policy analysis positions at different levels of government in the future.

Thirty-five students evaluated 128 programs, distributed across three years of PART summaries. The resulting data provides two decision points to compare student and OMB judgments, the PART program rating (where ineffective was given a score of one, results not demonstrated is two, adequate is three, moderately effective is four, effective is five), and a budget recommendation. Since 128 programs were evaluated, there are a total of 256 potential decision points.

Of the 125 valid program rating decision points, the students agreed with the OMB assessment 85 times, and disagreed 40 times. There was less agreement in terms of actual funding decisions. The raters only agreed with the OMB assessment on 65 of 127 occasions. The raters were somewhat more likely to recommend increased funding (on 40 occasions) than to cut it (on 22 occasions). Comparisons of correlations between PART evaluations and budget recommendations for the sample suggest that the decision behavior of students was similar to OMB officials. For both groups, PART ratings were positively correlated to proposed budget increases, indicating that programs that were rated as better performers tended to see receive higher scores. Students willing to disagree with the OMB on the program evaluation were also more likely to disagree on funding allocation. Disagreement on the evaluation and disagreement on the budget recommendation is positively correlated at .433, significant at .001 (two tailed).[li]

What conclusions, albeit tentative, can be reached? First, raters were generally reluctant to disagree with the ratings of the OMB. While the next section will discuss the reasons for disagreement, it is first worth considering the reasons for agreement. Students pointed to the information advantage enjoyed by the OMB, and the persuasiveness of the PART format, which allows the OMB to lay out detailed information to supports its recommendations. To be sure, the PART summaries are persuasive and unless someone is deeply interested in the program, and has the basis upon which to form a contrary opinion, they are likely to accept the OMB version.

Any experimental approach largely excludes consideration of institutional pressures that actual analysts face. The experiment therefore underestimates the likelihood of disagreement between parties because agency officials and legislative committee staff will not suffer from the same measure of information asymmetry as the students. They will also have more strongly-held preferences. We can expect that agency actors, stakeholders, and committee staff will be in a better position to launch plausible alternative interpretations to the PART summaries, and to be more motivated to do so. The experiment still provides some evidence that even individuals without particular institutional motivations are still capable of exploiting the ambiguity of performance information to produce alternative, but reasonable, assessments to those created by OMB analysts. In addition, because the experiment takes place among public affairs programs from two quite different schools, there are some limited insights to be gained on whether different institutional settings (being a student in one public affairs school or another) results in different decision outcomes. The likelihood that this is the case increases because of marked differences between the schools, even though the content of both public affairs programs is similar, and the context of the experiment was the same.

The students at the Bush School tend to be more conservative, more critical of government programs, and more likely to resist increases in federal government spending. Students at the La Follette School tend to be more politically liberal, and more supportive of government programs. The results for the experiment provide support for the hypothesis that institution matters to decision outcomes. Students from Wisconsin were more likely to disagree with the OMB, both in terms of PART ratings and budget proposals. A dummy variable indicating being a Wisconsin student was positively correlated with willingness to disagree with OMB PART ratings (at .325, significant at .001) and OMB budget assessments (correlated .288, significant at .01). Not only were Wisconsin students more likely to disagree, the nature of their disagreement was different to their Texas A&M counterparts and consistent with a more liberal perspective. Students at Wisconsin were significantly more likely to propose a higher rating relative to the OMB assessment (correlated at .321, significant at .001), and to be supportive of increased budgets for agencies (correlated at .381, significant at .001).

The Logic of Disagreement

The primary goal of the experiment is to undercover the logic that underpins disagreement about the same information. This helps inform how performance information is interpreted, and in turn, used. In general, students used the following rationales when disagreeing with the grades and the funding decisions:

- PART ratings were unreliable, usually based on a disagreement about the ability to measure the function, the role of external factors, the adequacy of the underlying performance data or an alternative interpretation of what the data meant.
- Programs with results not demonstrated, poor financial controls, or which have achieved their original goals should not receive increases in funding.
- Programs with strong positive assessments should receive higher funding.
- Programs should be assessed consistent with other PART programs and receive increases consistent with programs with similar PART ratings.
- Program problems, while real, are being eliminated through improved planning and management; program cuts are not appropriate.
- Clear relationship between resources, need and program delivery trumps performance information.
- Stakeholder and congressional views trumps performance information.

Due to length I exclude many specific examples that help illustrate student disagreements with PART, which can be found in the actual book and in a recent article.[lii] These examples suggest that the students employed reasonable warrants upon which to disagree with OMB conclusions about PART ratings and funding recommendations. The plausible nature of the student arguments is demonstrated by the fact that OMB examiners employed many of these same rationales – on what constituted a reasonable basis for assessing performance, on how tightly evaluations should connect to funding, on the importance of demonstrating results, and on the relative importance of program need and the connection of resources to service delivery. However, neither the OMB nor the students used these rationales in the same way for every decision, suggesting that how performance information is used depends partly on the performance information, partly on the decisionmakers involved, and partly on the nature of the decision to be made.

The experiment helps to shed light on the actual experience of PART. A GAO report has noted that “"OMB and agencies often have conflicting ideas about what to measure, how to measure it, and how to report program results”[liii]. Similar to the results of the experiment, GAO analysts offered a number of examples where agencies and congressional committee staff offered reasoned objections for disagreeing with the performance information chosen or excluded by the OMB, and with the PART assessments made.

Conclusion

This paper has presented dialogue theory to illustrate the ambiguity and subjectivity of performance information use. It is important to clarify that such ambiguity will not be equally present for all functions. In instances where programs are relatively simple with clear goals, costs and benefits, there is greater chance of avoiding ambiguity and generating agreement. Thurmaier and Willoughby draw on research in psychology to make the point that on some issues – particularly those perceived as simple, or where actors have familiarity or direct experience – actors are more likely to have strong exogenous preferences, and are less prone to persuasion.[liv] There is also greater potential for agreement in decision situations that feature a relatively homogenous set of decisionmakers and interests, because there is a reduced incentive for advocacy and a shared set of institutional norms.

However, the experience of PART illustrates just how prevalent ambiguity and subjectivity are in the process of generating and using performance information. In constructing PART, the OMB used its legitimacy and central position in the budget process to structure the discourse about performance. It has pushed agencies to develop the type of performance information it sees as important. Other parties are likely to resist PART assessments unless they align with their views. The basic ambiguity of PART assessments was demonstrated in the experiment described in this chapter. Students examining PART assessments were able to use logical warrants to come to different conclusions on what the underlying data meant for creating PART scores, and what the PART scores meant for budget decisions. Such disagreement will increase when the actors involved in the decision process represent particular roles and institutional interests, a point illustrated by the resistance by Congress and agencies to PART.

The insights of dialogue theory apply more broadly than the PART reform to encompass any situation where one party is trying to regulate the performance of another by using performance information. What does dialogue theory tell us about the decisions that arise from such regulatory relationships? There will be a variety of actors with different interpretations of performance data, making different arguments about the actions implied. The actors using the performance information have different positions, motivations, priorities and understanding of what the data means and what it actions should result from the information. In short, the use of performance information is a subjective exercise and rarely removes much of the ambiguity from decisions. Once we incorporate the assumption of roles and advocacy, we expect actors to selectively present performance data in a context that supports their point of view, and to discount conflicting information.

The political nature of decisionmaking will interact with, rather than be replaced by, performance information. Advocacy and ideology will continue to shape allocation decisions, using performance information as a tool. Rather than present information comprehensively, giving equal balance to all, actors will highlight specific pieces of data, offer plausible explanations for why performance occurs, and how it can be improved.

Performance information, when used, will not necessarily engender consensus and agreement. This depends greatly in the actors involved in the decision, their interpretation of the data, their ability to persuade others and their power in the decision process. In some cases, what one group of decisionmakers conclude is a reasonable interpretation and an appropriate response may be completely at odds with another group looking at similar information.

The nature of the performance information about a program is therefore not predictive as to the decision that will emerge about the program. The simplest illustration of this point is the basic ambiguity of choice between performance and resources. If a program is consistently performing well, does that indicate that it should receive greater resources, or that it is already amply provided for? Is the poorer performing program a candidate for elimination, of just in need of additional resources? Different actors might take the same set of performance data and offer plausible and logical arguments for either option.

References

[i] James G. March and Johan P. Olsen, *Ambiguity and Choice in Organizations*, (Bergen: Universitetsforlaget, 1976).
[ii] Martha S. Feldman, *Order Without Design: Information Production and Policy Making*, (Stanford, CA: Stanford University Press, 1989), 5.
[iii] March and Olsen, *Ambiguity and Choice in Organizations.*
[iv] Feldman, *Order Without Design*, 144.
[v] James G. March, "Ambiguity and Accounting: The Elusive Link Between Information and Decision Making," *Accounting, Organizations and Society* 12 (1987): 153-168, 160.
[vi] David Grant, Cynthia Hardy, Cliff Oswick and Linda Putnam (eds.), *The Sage Handbook of Organizational Discourse*, (London: Sage Publications, 2004).
[vii] Ann B. Blalock and Burt Barnow, "Is the New Obsession with Performance Management Making the Truth About Social Programs," in Dall Forsythe, ed. *Quicker, Better, Cheaper?: Managing Performance in American Government*, (Albany, NY: The Rockefeller Institute, 2001), 485-517.
[viii] Stephen Toulmin, *Uses of Argument*, (London: Cambridge University Press, 1958).
[ix] March and Olsen, *Ambiguity and Choice in Organizations.*
[x] John Kingdon, *Agenda, Alternatives and Public Policies*, (Boston: Little, Brown, 1984)
[xi] March, "Ambiguity and Accounting," 154.
[xii] Barbara Stone, *Policy Paradox: The Art of Political Decisionmaking*, (New York: W.W. Norton and Company, 1997), 167.
[xiii] Stone, *Policy Paradox*, 177.
[xiv] Carolyn Heinrich and Elizabeth Fournier, "Dimensions of Publicness and Performance in Substance Abuse treatment Programs," *Journal of Policy Analysis and Management* 23 (2004): 49-70, 63.
[xv] Stone, *Policy Paradox*, 169-170.
[xvi] Raymond Nickerson, Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, *Review of General Psychology*, 2(1998): 175-220.
[xvii] Stone, *Policy Paradox*, 187.
[xviii] Nelson Phillips, Thomas B. Lawrence and Cynthia Hardy, "Discourse and Institutions," *Academy of Management Review*. 29 (2004): 635-652.
[xix] Kurt Thurmaier and Katherine Willoughby. *Policy and Politics in State Budgeting.* (New York: ME Sharpe, 2001), 83.
[xx] Aaron A. Wildavsky and Naomi Caiden, *The New Politics of the Budgetary Process*, (5 ed.) (New York: Longman, 2003).
[xxi] Julia E. Melkers and Katherine Willoughby, "Staying the Course: The User of Performance Measurement in State Government." Washington DC: IBM Center for the Business of Government, 2004.
[xxii] The team also performed the role of determining what the official OMB position would be on a number of questions the exercise would raise, e.g. how do PART scores related to budgeting? What is the relationship between PART and GPRA? The recommendations of the team can be found at: http://www.whitehouse.gov/omb/mgmt-gpra/pmac_part_revisions.pdf
[xxiii] In an attempt to modify the tool for different types of programs (direct federal programs, competitive grant programs, block/formula grant programs, regulatory based programs, capital assets and service acquisition programs, credit programs, and research and development programs) the OMB created seven versions of PART, with each version having approximately 30 questions.
[xxiv] OMB examiners may adopt an alternative weighting of questions if they consider appropriate, see U.S. Office of Management and Budget, *Budget Procedures Memorandum No. 861*, (Available at http://www.whitehouse.gov/omb/budget/fy2005/pdf/bpm861.pdf, 2003), 15.
[xxv] U.S. OMB, *Budget Procedures Memorandum No. 861*.
[xxvi] U.S. Government Accountability Office, *Performance Budgeting: PART Focuses Attention on Program Performance, but More Can be Done to Engage Congress* (Washington DC: GAO, 2005), 17.
[xxvii] Harry Hatry, *Comments of the Members of the Performance Advisory Council*, (Available at http://www.whitehouse.gov/omb/budintegration/pmac_030303comments.pdf , 2003), 8.
[xxviii] U.S. Government Accountability Office, *Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget* (Washington, D.C.: GAO, 2004), 25.

[xxix] Ibid, 24.
[xxx] Clinton Brass, *The Bush Administration's Program Assessment Rating Tool,* (Washington, D.C.: The Congressional Research Service, 2004), 21.
[xxxi] US Government Accountability Office, *Observations on the Use of OMB's Program Assessment Rating Tool.*
[xxxii] Ibid, 21.
[xxxiii] Ibid, 7.
[xxxiv] Cynthia Hardy and Nelson Philips, "Discourse and Power," In David Grant, Cynthia Hardy, Cliff Oswick, Linda Putnam (eds.), *The Sage Handbook of Organizational Discourse,* (London: Sage Publications, 2004), 299-316.
[xxxv] Nelson Phillips, Thomas B. Lawrence and Cynthia Hardy, "Discourse and Institutions," *Academy of Management Review* 29 (2004): 635-52.
[xxxvi] Aaron A. Wildavsky, "The Political Economy of Efficiency: Cost-Benefit Analysis, System Analysis, and Program Budgeting," *Public Administration Review* 26 (1968): 292-310, 308.
[xxxvii] Beryl Radin, Testimony to the Senate Homeland Security and Government Affairs Subcommittee On Federal Financial Management, Government Information, and International Security, June 14 2005.
[xxxviii] Wildavsky, "The Political Economy of Efficiency, 302
[xxxix] U.S. Government Accountability Office, *Performance Budgeting: PART Focuses Attention on Program Performance,* 7-8, 37-38.
[xl] US Government Accountability Office, *Observations on the Use of OMB's Program Assessment Rating Tool*, 23.
[xli] U.S. Government Accountability Office, *Program Evaluation*
[xlii] Ibid, 16-17.
[xliii] Barry White. "Examining Budgets for Chief Executives," in Roy Meyers (ed), *Handbook of Government Budgeting*, (San Fransicso: Jossey-Bass, 1999), 462-484; Bernard H. Martin, Joseph T. Wholey and Roy T. Meyers, "The New Equation at the OMB: M+B=RMO," *Public Budgeting and Finance* 15 (1995): 86-96.
[xliv] Ibid.
[xlv] Ibid, 22.
[xlvi] This observation has been made anecdotally, and based on interviews with OMB and congressional staff. It received further support when the GAO noted that "…it is not clear that the PART has had any significant impact on congressional authorization, appropriations, and oversight activities to date. Moreover, it is unlikely that performance information will be used unless it is believed to be credible and reliable and reflects a consensus about performance goals among a community of interested parties." U.S. Government Accountability Office, *Performance Budgeting: PART Focuses Attention on Program Performance*, 44.
[xlvii] U.S. Government Accountability Office, *Performance Budgeting: PART Focuses Attention on Program Performance*, 8, 42-46.
[xlviii] Matthew D. McCubbins and Thomas Schwartz, "Congressional Oversight Overlooked: Police Patrol versus Fire Alarms," *American Journal of Political Science,* 28(1984): 165-179. The OMB has made access to information easier though a more user-friendly website that provide PART analyses: www.exepctmore.gov.
[xlix] Students were required to provide a written memo of about 7 single-spaced pages explaining their decisions. They also had to present their findings before the class, who operated as a mock legislative committee, asking pointed questions about the recommendations. After one session, a group of eight students served as a focus group to discuss in greater detail the reasons for the decisions they made.

[l] Bretschneider, Stu, Jeffery Straussman and Dan Mullins "Do Revenue Forecasts Influence Budget Setting? A Small Group Experiment," *Policy Sciences* 21 (1988): 305-325; Coursey, David H., "Information Credibility and Choosing Policy Alternatives: An Experimental Test of Cognitive-Response Theory" *Journal of Public Administration Research and Theory*, 2 (1992): 315-331; Wittmer, Dennis, "Ethical Sensitivity and Managerial Decisionmaking: An Experiment," *Journal of Public Administration Research and Theory*, 2(1992): 443-462. A argument for the use of experiments in public policy and management, and a review of relevant work can be found in Bozeman, Barry and Patrick Scott. "Laboratory Experiments in Public Policy and Management," *Journal of Public Administration Research and Theory*, 2(1992): 293-313. A more general argument for the use of experiments in assessing psychological behavior can be found in Mook, Douglas G. "In defense of external invalidity," *American Psychologist*, 38 (1983): 379-387.

[li] In order to control for the role of outliers I converted the variables into a simplified disagreement scale, where students evaluating the program higher than the OMB is scored at -1, evaluating the program lower is 1, no

disagreement is 0; students recommending higher budgets than the OMB is -1, recommending a lower score is 1, no disagreement is zero.

[lii] Moynihan, Donald P. "What Do We Talk About When We Talk About Performance: Dialogue Theory and Performance Budgeting." *Journal of Public Administration Research and Theory* 16(2): 151-168.

[liii] U.S. Government Accountability Office, *Performance Budgeting: PART Focuses Attention on Program Performance*, 7.

[liv] Thurmaier and Willoughby, *Policy and Politics in State Budgeting.*