

The Role of Program Evaluation in Environmental Policy: A Critical Evaluation of Incentive Structures

Lori S. Benneer
Nicholas School of the Environment and Earth Sciences
Duke University
Box 90328
Durham, NC 27708
(919) 613-8083
lori.benneer@duke.edu

Katherine L. Dickinson
Nicholas School of the Environment and Earth Sciences
Duke University
Box 90328
Durham, NC 27708
katie.dickinson@duke.edu

The Role of Program Evaluation in Environmental Policy: A Critical Evaluation of Incentives Created by GPRA and PART

Lori S. Benneer and Katherine L. Dickinson*

1. Introduction

While there is legitimate debate on the optimal role and size of government, most everyone agrees that government agencies should be accountable to achieve results and improve efficiency of service provision over time. That is, government should be performance-focused. Despite the obvious benefits of a performance-focused system, many government programs are not systematically evaluated for performance. Rather many government programs continue to be evaluated based on completion of processes rather than on attainment of performance (Metzenbaum 1998, Benneer and Coglianesse 2004).

While these problems exist throughout government, our focus is on the use of evaluation in environmental policy. Our interest in performance-focused environmental policy is not entirely selfless. We actively work in the area of program evaluation for environmental policies and therefore have a professional interest in understanding how these evaluations are used within government. However, there are other important reasons to target environmental policy and the use of evaluation in this area. In particular, there appears to be a disconnect between what would pass for performance-

* Benneer is an Assistant Professor of Environmental Economics and Policy at the Nicholas School of the Environment and Earth Sciences, Duke University. She can be reached at (919) 613-8083 or lori.benneer@duke.edu. Dickinson is a Ph.D. candidate in Environmental Economics and Policy at Duke University. She can be reached at katie.dickinson@duke.edu.

based evaluation in academic circles and what would pass for performance-based evaluation on the ground within the Environmental Protection Agency. In part this is a perceived disconnect, based on anecdotal evidence from interactions with state and federal agency personnel through science advisory panels, conferences, and as researchers. But there is some evidence that EPA lags behind other federal agencies in conducting and using program evaluation. A 2000 report by the Government Accountability Office highlights difficulties in measuring results and developing a performance focus at the United States Environmental Protection Agency (EPA). The purpose of this research is to identify the barriers to performance-focused environmental policy and critically evaluate the likelihood that current policy initiatives will help EPA overcome these barriers and move the agency toward a more performance-based system.

This paper begins by describing an “ideal” model of performance-focused environmental policy and outlines the barriers to establishing such a performance-based system. We then examine two recent initiatives designed to move agencies toward a performance-focused approach. One initiative is legislative, the Government Performance and Results Act (GPRA), and one is administrative, the Program Assessment Ratings Tool (PART). This paper asks three questions: 1) How do GPRA and PART address the barriers to agencies’ use of evaluations in order to improve environmental performance? 2) How are agencies likely to respond to the set of incentives these initiatives create? And finally, 3) how can these initiatives be improved to move government closer to the performance-focused ideal?

The results of our analysis suggest that the incentives created by GPRA and PART will not ensure that all agencies conduct evaluations and use these evaluations to

improve performance. In particular, the fact that performance cannot be measured with certainty, but rather is subject both to external shocks and to manipulation of the performance metric, implies that agencies may respond strategically to GPRA and PART incentives in ways that do not enhance performance. While under some circumstances, the optimal strategy will be to learn from evaluations and invest in improved performance, in other circumstances it may make more sense to manipulate the performance metric and manage to the performance target. We argue that these perverse incentives can be mitigated by the use of third-party reviews and by an increased focus on building social and organizational cultures that support and maintain the ideals of a performance-based system.

2. Ideal Model of Policy Design and Evaluation

Historically, many environmental agencies have been concerned largely with the *processes* of environmental management—developing regulations, writing permits, conducting inspections, issuing compliance orders and fines, encouraging participation in voluntary programs, and so forth (Metzenbaum 1998, Bennear and Coglianesi 2004, GAO 2000). While these processes certainly have some relation to improving environmental performance, these relationships are imperfect and often poorly understood. Moreover, evaluation of the agency's activities has been based on whether the agency has complied with administrative processes rather than on the actual performance in attaining its mandate. This process-focused approach has led to “bean counting” whereby agency efficacy is determined by the number of inspections, the number of fines, the number of “beans” (Metzenbaum 1998, GAO 2000). In contrast, a performance-focused system is one in which agencies are judged by the effect of their activities on the ultimate policy outcomes of interest.

Performance-focused policy can be viewed as a cyclical process involving three key steps: learning, action, and evaluation. Evaluation then marks the end of one cycle of policy design, but it can also mark the beginning of the next cycle if the results of the evaluation are used to learn about the effectiveness of different policy options and inform future choices (Brewer and deLeon 1986, Bennear and Coglianese 2004). This process is diagrammed in Figure 1. This section describes the ideal learn-act-evaluate cycle in more detail.

Learning is at the heart of a performance-based system. A focus on “performance” is only possible to the extent that we know what performance is, how to measure it, and how to assimilate performance-related information in a meaningful way. Learning is thus essential to a performance-based approach. In the context of government agency decisionmaking, we envision “learning” as a three-step process involving: 1) gathering information relevant to the decision; 2) integrating the information (e.g., deciding what information is valid, weighting aspects or attributes, making tradeoffs); and 3) assessing the information in order to determine the best course of action.

The ideal model implies certain behaviors at each stage of this process. Information gathering should be thorough and efficient, covering all relevant information sources and ignoring irrelevant data. In the integration stage, decision makers should objectively judge the validity of information. They must then weight different kinds of information based on some stable set of social preferences, making tradeoffs between competing objectives (e.g., minimize costs, maximize safety). Finally, decisionmakers should assess different options and apply a consistent rule (e.g., expected social welfare maximization) to decide among different policies or approaches.

Based on this definition, learning is a fairly intangible process. In particular, learning (or knowing what to do) is separate from action (doing it), and therefore, there may be no observable record of the learning process. Instead learning increases the range of potential policy options, which increases the organization's ability to respond to changing circumstances over time (Cohen & Sproull year???)

What we do observe is the next step in the policy process: action. In the ideal model, this step is straightforward. In the learning stage, the decision maker identifies the optimal policy, and in the action stage, she simply implements that policy. In practice, of course, this step is more complicated. Even when a decision maker knows what she should do or would like to do, a number of time, logistical, budgetary, administrative, and other constraints may prevent her from doing it.

These constraints, and the disconnect between "knowing what to do" and "doing it" is well illustrated in a recent book by Bazerman and Watkins (2004). These authors define a "predictable surprise" as "an event or set of events that take an individual or group by surprise, despite prior awareness of all of the information necessary to anticipate the events and their consequences." They argue that the terrorist attacks of September 11 and the collapse of Enron represent two examples of contemporary, high profile predictable surprises. The recent catastrophe resulting from Hurricane Katrina also conforms to this definition. In all of these cases, groups of individuals within government recognized that a problem existed, learned about appropriate solutions, and even made recommendations for addressing the problem, but these recommendations were never implemented. These examples illustrate the fact that the real world often deviates from our ideal model in which action follows smoothly and optimally from the learning process.

The next step in the policy cycle is evaluation. After an appropriate amount of time, the next step in a performance-based system is to evaluate the results of a particular course of action. The purpose of this evaluation is to provide a measure of the policy's outcomes. Did the policy have the intended results? Were there any unforeseen or unintended consequences? In retrospect, how well did this policy perform? Answers to these questions provide information about the effectiveness of the particular approach that was chosen in a specific case, and can serve as inputs into future learning and decision making (Bennear and Coglianesse, 2004).

In the ideal model, evaluation of a particular policy or program should have several important characteristics. First, the evaluation should measure the policy's impacts on a range of relevant outcomes, where "relevant" is defined to cover all important stakeholders' interests. Second, the evaluation should attempt to separate the effects of the policy under study from other, simultaneously occurring events. This usually involves constructing some counterfactual account of what would have happened in the absence of the policy. Third, the evaluation should employ methods appropriate to the relevant questions and available data. Some questions are best answered through quantitative, econometric studies (e.g., What were the measurable impacts of this policy on variable X?), while others require more qualitative approaches (e.g., Why did this policy succeed or fail? What were stakeholders' attitudes toward this approach?). Systematic, objective evaluations of this kind connect action and learning, providing a fundamental feedback loop that is essential to a performance-based system.

Not surprisingly, actual government decision making often deviates from the ideal performance-based paradigm outlined above. In this paper, we outline some of the possible reasons why this might be the case. We provide a conceptual framework

that identifies the cognitive, social, structural, and incentive barriers to achieving an optimal performance-based system. Since the purpose of this paper is to examine how government can improve the use of evaluations in policy design, we tend to focus on barriers to the use of evaluations in the learning process and on barriers to conducting evaluations after programs have been implemented.

3. Barriers to Implementing the Ideal Model

Barriers to implementing the ideal policy model within government agencies can be grouped into four broad categories: cognitive barriers, social barriers, imperfect organizational structure, and imperfect incentives. There is extensive literature drawing on psychology, economics, management, sociology, and political science on these four categories of barriers and we cannot do justice to all of these literatures in this summary review. Instead we highlight the key insights from each of these fields in order to illustrate how an oversight body, such as Congress or the White House, might be able to address these barriers. In particular, we will use the insights from these fields to frame our analysis of the likely effects of two specific congressional and White House programs, namely GPRA and PART, on evaluation and performance within government agencies.

3.1. Cognitive Barriers

Recent work in behavioral decision theory highlights some barriers to the ideal performance-focused model, particularly affecting the learning stage. This literature provides key insights in assessing what kinds of information people are likely to use in their decision making processes, how they weight and integrate that information, and how this information influences the ultimate decision.

One cognitive barrier to the effective use of evaluations during the learning part of the policy process is the reliance on various heuristics for decision making. Rather than collecting, processing, and evaluating all of the available information before making a decision about what action to take, people (and hence agencies composed of people) may rely on “rules of thumb” or shortcuts in the learning process (Tversky & Kahneman 1974).

One of the most relevant heuristics in the context of our ideal learn-act-evaluate model is the availability heuristic. This shortcut consists of judging the likelihood of an uncertain event by the ease with which instances of similar events can be brought to mind. In the context of government agency decision making, the availability heuristic may have a strong influence on the information gathering and integration steps, determining what evidence is most relevant for the task at hand and giving more weight to certain kinds of information. This can lead to biases if, for example, the availability heuristic causes only very recent experience or certain kinds of studies (e.g., descriptive case studies or in-house evaluations) to be considered.

Another important and well-documented tendency is “confirmation bias” or the tendency of people to seek information that confirms their prior beliefs about the correct action (Klayman, 1995). For example, an agency that is already convinced that policy X reduces pollution may look for evidence that policy X indeed decreased pollution in some specific case and discount any evidence that suggests that policy X did not decrease pollution. This may lead to less than optimal use of information from prior evaluations in program design. It may also bias the results of evaluations themselves if these evaluations are conducted “internally” by individuals with a stake in the evaluations’ results.

The final set of cognitive barriers concerns how agencies integrate and assess information in decision making. Research suggests that people may be biased toward maintaining the status quo because the strain in weighing the different attributes (both positive and negative) of a decision to change away from the status quo is demanding. This can lead to decision avoidance (Lee 1998, Anderson 2003). Similarly, research has shown that people tend to prefer errors of omission (doing nothing) to errors of commission (doing the wrong thing) which may also lead to a bias toward the status quo (Spranca et al 1991).

3.2. Social Barriers

While the focus in the previous section was primarily on the individual, in the context of a performance-based system we are really interested in the ability of an organization (i.e., a government agency) to engage in successful learning, action, and evaluation. This section explores how interactions between individuals within an organization may produce additional barriers to learning.

The organizational learning literature emphasizes several ways in which membership in an organization shapes individual learning. Perhaps the most important factor involves the way in which an organization influences an individual's identity. Being a part of an organization creates a sense of shared purpose and a collective set of norms, which influence how an individual sees herself, what goals she pursues, and how she works to achieve those goals. Through the organization, individuals become part of a "community-of-practice" (Brown & Duguid xxxx), interacting and learning from each other and adopting a common viewpoint. These shared norms and common identities serve an important purpose within the organization, facilitating coordination and communication (Kogut & Zander 1996). However, an individual's identity also

limits the way he thinks, putting constraints on what options he considers as well as how he searches for information (Kogut & Zander 1996). For example, people may experience a kind of cognitive dissonance (Kogut & Zander 1996) when faced with evidence that does not conform to their identity. In the case of a government agency, an evaluation or other piece of evidence showing that some previous policy was not successful may conflict with individuals' image of themselves and of the agency in general. Confronted with this cognitive dissonance, individuals may discard certain kinds of information or refuse to consider certain interpretations of the available evidence.

More generally, identity can be a powerful factor affecting learning, action, and evaluation within an organization. For example, Powell's (2000) study of the use of science at the EPA found that EPA's identity as a regulatory agency often "subordinates" science within the agency. Similarly, Mendeloff's (2004) comparison of the role of evaluation at the Occupational Safety and Health Administration (OSHA) and the National Highway Transportation Safety Administration (NHTSA) revealed that OSHA's identity as an enforcement agency contributes to a lack of systematic evaluation, while NHTSA sees itself as a "science-based regulatory agency" and relies more heavily on evaluation. In this case, OSHA's stakeholders play a strong role in reinforcing this identity. The firms and industries OSHA regulates are hostile to evaluation, contributing to an atmosphere where this part of the learning process is not seen as important (Mendeloff 2004). These studies are consistent with a study by the Government Accountability Office (2003) which emphasized the existence of an "evaluation culture" as a key factor driving creation and use of evaluation in the

agencies this study identified as success stories. This “institutional commitment to learning from evaluation” (GAO 2003) is an important part of these agencies’ identities.

3.3. Barriers in Organizational Structure

The organizational chart is often one of the first things to be examined when considering the performance of any organization. How is the organization structured and how does this structure facilitate or impede the organization’s goals? For evidence of the primary role organization structure plays in a critique of organizational performance one need look no further than the U.S. experience following the terrorist attacks on September 11, 2001. Following what by all accounts was a massive performance failure among U.S. intelligence agencies, the immediate proposals to improve performance in this sector focused on restructuring the organizational hierarchy, in particular the creation of the Department of Homeland Security and the placement of intelligence agencies within that Department (cite ???).

This focus on the organizational chart is not surprising. An organization’s structure can impede or facilitate communications, coordination, evaluation and learning. Overly vertical agency structure may result in a series of “fiefdoms” each with its own goals, and projects. Because communication must flow up the hierarchy before it can be passed across the hierarchy, learning from others’ programs and evaluations is limited. In contrast, overly horizontal structures might create “decision by committee,” a situation where everybody is responsible for decisions, and hence, nobody is responsible for decisions. This lack of direct accountability can also impair the policy process.

There is evidence that organizational structure can facilitate or impair agencies’ ability to conduct evaluations and to use these evaluations in policy design. The GAO (2003)

studied five agencies that were viewed as having a strong performance focus. They found that the ability to develop collaborative partnerships with other agencies and across federal, state and local levels played a key role in their success. As a more specific example, Mendeloff (2004) cites organizational structure as one of the key barriers to performance-based approaches at the Occupational Safety and Health Administration (OSHA). In particular, Mendeloff highlights the fact that the National Institute of Occupational Safety and Health (NIOSH), which is the primary agency responsible for research on occupational health and safety issues, is not located within OSHA, but rather within the Bureau of Labor Statistics. This feature of OSHA's organizational structure inhibits this agency's ability to conduct evaluations that respond to its information needs and encourage learning and improved performance over time.

There is another organizational barrier to the use of evaluation in policy design which has to do with the level of existing expertise in evaluation methods within different agencies. In 1998, the GAO surveyed 23 different government offices that had conducted some program evaluation during 1995. They found that half of these offices had fewer than 18 full time-equivalent employees (FTEs), and the level of resources spent on evaluation was also low (GAO 1998). Similarly, Mendeloff (2004) argued that one of the reasons NHTSA performs more evaluations than OSHA is that NHTSA has greater institutional capacity for evaluation activity.

3.4. Incentive Barriers

A final set of barriers to the ideal model of policy design concerns the differences in the goals and objectives of government agencies, on the one hand, and the organizations that oversee them (such as Congress and the White House), on the other. One can think of the government agencies acting as agents on behalf of a principal. In

the case of Federal agencies, the principal is the citizenry, but they are generally represented by the Congress or the White House. Principal-agent models of this kind involve a fundamental conflict between the objectives of the two actors, the principal and the agent. In particular, the agent may not find it in her best interest to work toward the principal's goals (Grossman and Hart 1983, Hart and Holmstrom 1987).

In the case of government agency decision making, the principal (Congress or the White House) wants government agencies to work toward one set of objectives and achieve certain performance-based results outlined in the ideal model. However, there are a number of reasons why the agencies and the people who work in them may not find it in their interest to pursue the principal's goals. That raises the question: What are the agencies' goals? Traditionally we think individuals are motivated by salary. But in thinking about organizations, the objective is less clear.

Unfortunately, there is no single, well-articulated objective that applies to all government agencies and programs. However, agencies (or program managers within agencies) are often thought to be concerned with maximizing their budget (Niskanen 1994). The agency's goal of maximizing its budget may not be consistent with the principal's goal of increasing efficiency and improving performance. Being very efficient in use of resources by carefully choosing programs and policies, evaluating those policies and making changes, may not be the best way for an agency to increase its budget. At the extreme, if an agency is very efficient at achieving its mission, the urgency of that mission may be diminished. It is not much of a stretch to imagine members of Congress saying, "We used to have a problem with dirty water, but EPA was so efficient we no longer have a problem. Let's give their funding to Homeland Security." Thus, the challenge is to create a set of incentives for agencies to gather

relevant information, use it in assessing policy options, and then provide incentives on the back end for agencies to engage in further evaluation and use the results of those evaluations to improve policy design.

In summary, there are number of different cognitive, social, organizational, and incentive-related barriers to achieving the optimal performance-based policy system. These barriers are summarized in Table 1. The remainder of this paper examines two recent attempts by Congress and the White House to address these barriers, with particular focus on changing the incentives facing government agencies. We analyze these new incentives in light of the full range of barriers outlined above.

4. GPRA and PART

During the 1990s there was a push toward institutionalizing a performance-based governance system by requiring agencies to set performance goals, regularly assess performance, and use the results of these assessments to improve programs. In 1993, Congress passed the Government Performance and Results Act (GPRA). The act was motivated by a desire to improve government performance, budgeting and program oversight by requiring federal programs to set performance-based goals, measure progress toward those goals, and publicly report this progress (U.S. Code, ???).

There are three primary components of GPRA requirements. First, agencies are required to engage in strategic planning. GPRA requires that agencies' strategic plans identify performance goals, describe how these goals will be achieved, and describe methods used for evaluating progress towards these goals. Whenever possible, agencies should use outcome-based performance measures in stating goals—that is, agencies should use measures that are directly related to the intended purpose of the program

(e.g., improvements in water quality) and avoid measures that detect the output of the program (e.g., number of permits issued).

The second component requires annual performance reports. These annual reports should provide evidence of progress towards the goals outlined in the strategic plan, re-evaluate the plan based on evidence of past performance, and provide rationale for any programs that have not met their performance goals.

The final component of the GPRA legislation allows agencies to waive some administrative procedural requirements in exchange for committing to a performance-based system. Agencies that are willing to commit to performance accountability can waive mandates on staffing levels, pay levels, and restrictions on funding transfers among programs. The purpose of the administrative procedure waiver is to allow agencies more flexibility to achieve greater performance results.

The GPRA requirements are phased in over seven years. The requirements are first applied to pilot programs in 1994, 1995, and 1996 with broad applicability in 1997. The procedural flexibility component was tested on pilot project in 1995 and 1996 with wider availability in 1999.

While GPRA sets out a framework for performance-based planning and evaluation, the legislation is short on specific requirements. For example, while GPRA requires agencies to report on progress towards their performance goals, the legislation provides no guidance on what constitutes acceptable evidence of progress. Similarly, while performance reports are required, there is no discussion of what might occur if performance is not demonstrated. In particular, the GPRA legislation does not directly tie performance reports to the budgeting process.

Within the executive branch, the Office of Management and Budget (OMB) also began a pilot program to promote performance-based evaluation of government projects. The Program Assessment Rating Tool (PART) was developed to assist OMB in evaluating government programs and to link program performance to budget decisions. The PART complements GPRA – many of the requirements of GPRA are also present in the PART assessment, but PART also goes beyond GPRA requirements.

There are four components of the PART: program purpose and design, strategic planning, program management, and program results and accountability. Each part contains a series of yes/no questions. Each question is assigned a weight and the total score reflects the weighted average of scores where a yes receives a point value of 1 and a no a point value of zero. For example, in the program purpose or design section there are five questions focused on the mission of the program, its clarity and potential duplication with other programs. The other sections are organized in a similar fashion. The strategic planning section contains eight questions that focus on short and long term goals, the existence of performance measures and baselines from which improvements in these measures can be judged, the existence of independent evaluation of program progress and the connection between such evaluations and the budgeting and strategic planning process. The management section contains seven questions regarding the collection of information within the program, management of contractors and program partners, and management of funds. The section on program results and accountability contains five questions on whether the program has attained its short-term goals and is

making progress toward long-term goals. In addition, there are several questions that pertain only to some programs.¹

By default, all questions within a given section receive equal weight. However, the user is allowed to change the weights as appropriate for a particular project. For each question, detailed guidance is provided on what constitutes a “yes” response, often with specific examples drawn from actual programs.² For programs that have well-defined performance metrics, the overall score results in a ranking of effective (85-100), moderately effective (70-84), adequate (50-69) and ineffective (49 or less). Programs without well-defined metrics are assigned a rating of “results not demonstrated.”

To date, 399 programs have participated in the PART assessment. OMB selected programs as pilot PART programs in consultation with the agencies. The initial drafts of the evaluations are generally conducted by agency staff with OMB oversight. The final drafts are produced by OMB.³

How do GPRA and PART address the four categories of barriers to performance focused environmental policy? There appears to be little in the two programs that directly respond to cognitive and social barriers. There is one question on the PART form that addresses whether independent evaluations of the program have been conducted. At some level, requiring independent evaluations helps overcome some of the cognitive and social barriers—in particular, requiring third party evaluation should remove some of the confirmation bias. However, it appears that these two programs

¹ The PART divides programs into seven categories: (1) Direct Federal Program, (2) Competitive Grant Programs, (3) Block/Formula Grant Programs, (4) Regulatory Based Programs, (5) Capital Assets and Service Acquisition Programs, (6) Credit Programs, and (7) Research and Development Programs. Additional questions are specific to each program type.

² A complete list questions and guidance is available at <http://www.whitehouse.gov/omb/part>.

³ See http://www.whitehouse.gov/omb/part/2004_faq.html for more details.

generally do not focus on cognitive and social barriers. This is to be expected to some degree because these barriers are the most difficult to overcome by legislating action from the outside. While one can legislate or require agencies to conduct assessments, one cannot as easily create a culture where such evaluation is truly valued as an important part of the policy process. Addressing cognitive barriers operating at the individual level may be even more challenging.

Similarly, there is little focus on organizational barriers. PART does have some questions that address organizational problems, including whether the program is duplicative of other programs and whether there is coordination between all relevant agencies and programs. Indeed, one might think of the PART program itself as being a response to organizational barriers as the PART program increases the oversight and advisory role of OMB in the policy design process across agencies. GPRA also allows for relief from some organizational constraints if agencies are willing to commit to performance evaluation, but there are few specifics on how these changes will be facilitated.

Given their minimal impact on cognitive, social, and organizational barriers, the primary barrier that GPRA and PART seem to address is one of incentives. The two programs address incentives in slightly different ways. GPRA's primary incentive mechanism is through managerial flexibility. Agencies or individuals that agree to be held accountable to specific performance goals can waive normal restrictions on compensation (for individuals) and budget transfers (for organizations). Under the PART, the primary incentive comes through budget allocations. However, the relationship between a program's PART assessment and its budget allocation is not

perfectly monotonic. That is, receiving a low PART score does not necessarily result in budget cuts (OMB 2005).

Nonetheless, there is a correlation between PART scores and percentage budget changes. Table 2 shows the relationship between percent change in program budget and PART rankings for different programs over the fiscal years 2003-2004 and 2004-2005. From 2003 to 2004, budgets for all PART programs rose by an average of 4.2 percent. However, budgets for “effective” and “moderately effective” programs rose by more, and budgets for “adequate” and “results not demonstrated” programs rose by less. “Ineffective” programs saw their budgets decline by an average of 4.2%. From 2004 to 2005, there was an average decline in budgets of 1.2 percent. However, “effective” and “moderately effective” programs had budget increases while “adequate,” “ineffective,” and “results not demonstrated” programs saw larger than average decreases. Thus, while PART results do not determine budgetary outcomes, these findings demonstrate a positive correlation between PART rankings and subsequent budget changes.

One might argue that by changing the incentive structures, GPRA and PART are also addressing the other three barriers, albeit indirectly. If demonstrating continual performance improvements will increase agencies’ budgets, then perhaps the agency will work harder to overcome any cognitive, social or organizational barriers to learning, action, and evaluation. This seems at least plausible. What remains is to examine how we might expect programs to respond to the incentives created by GPRA and PART. That is the focus of the remainder of this paper.

5. A Model of Agency Incentives Under GPRA and PART

In order to assess how programs may respond to the incentives created by GPRA and PART, we build a simple two-period model. In the first period, an agency receives a

budget B_1 and a performance target P^* . At the end of the year an evaluation is conducted to reveal the agency's performance (P_t), which is some function of its budget, $P_t=Y_t(B_t)$. Agencies whose performance exceeds their target receive a budget bonus of β in the next period. Other agencies do not receive the bonus. Thus, the budget is given by $B_{t+1}(P_t, P^*)=B + \beta I(P_t \geq P^*)$ where β is the budget bonus and $I(P_t \geq P^*)$ is an indicator function that takes a value of 1 if the condition inside the parentheses holds and 0 otherwise.

In the second period, agencies can choose to invest in performance. Investment costs the agency some amount, C , but investment in innovation yields a new production function $Y(B)$ that gives higher performance for every budget level. One can view this type of investment as agencies using resources to learn from evaluation, making modifications to the program, and so forth.

While performance will improve with investment in innovation, agencies do not know in advance how much performance will be improved. The key point is that agencies will not know for sure that investment in innovation will raise their performance above the performance target. At the end of the period, the agency is again evaluated and performance bonuses are awarded to agencies that exceed their performance targets. Figure 1 represents the decision tree for each agency facing this set of incentives.

Even in this simplest formulation, where performance targets are fixed and there is a lump sum positive budget bonus for exceeding the target, the model reveals some interesting insights. First, the budget bonus does not provide sufficient incentive for continual investment in performance improvements. Agencies that exceed their performance target in period 1 have no incentive to improve performance further. For

agencies that fail to meet the performance target in period 1, the decision to invest in performance will be based on the expected costs and benefits of investment. The cost of investment is known in advance (C), but expected benefits depend on the magnitude of the budget bonus as well as the (subjective) probability that investment will lead to performance above the target. If the expected reward is greater than the cost, agencies will invest, otherwise they will not.

Given these incentives, innovation is likely when agencies are fairly close to the performance target, so that a small change in performance would guarantee the bonus. Innovation is also likely when investment in improved performance is not particularly costly. This might be the case if the agency was very inefficient in its use of resources initially, so that diverting labor and management attention to performance improvement does not create much “real” cost.

While this model is highly simplified, it helps clarify how incentives for increased performance might be institutionalized. In particular, the key drivers of incentives in this model are: (1) the performance target, (2) the performance metric, and (3) the budget bonus for good performance. These are the three features that GPRA and PART focus on as well. In the following sections we analyze how altering each of these components is likely to affect evaluation and performance.

5.1. Changing Performance Targets

GPRA allows agencies to set their own performance targets. There is no guidance in the law on how these targets should be determined. Under GPRA, one might be concerned about the lack of incentives for performance improvement above the performance target. If agencies are at liberty to choose their own performance targets, why not choose targets that are attainable? If the targets are attainable, then the model

suggests there is little incentive for better use of information gleaned from evaluations to improve program performance. In the language of the model, there is no incentive for investment in innovation.

The PART program is more specific regarding performance targets. Agencies must develop “ambitious targets” for both long-term and annual performance metrics. In addition, the agency must include at least one efficiency metric and therefore have at least one ambitious efficiency goal. Efficiency metrics measure how much “performance” an agency attains for a given amount of resources. The requirement for an efficiency goal can be seen as creating a performance target that is continually more stringent in each period. The requirement to have ambitious long-term as well as annual goals also acts as a way to continually ramp up the performance target. It is not sufficient for agencies to continue to attain the same performance every year, but rather agencies must be working each year toward a longer-term target of improved performance.

How do we expect agencies to respond if, rather than attaining a fixed target, they must continually improve their performance in each period? The model can be adjusted to encourage continued improvements in a variety of ways. One way is to change the performance target so that rather than receiving a bonus if performance exceeds some fixed target, the agency receives a bonus if performance in the current period exceeds performance in the previous period by some specified amount. The

budget in future periods is now given by: $B_{t+1}(P_t, P_{t-1}) = B + \mathcal{I}(P_t \geq \alpha P_{t-1})$ where $\alpha \geq 1$.⁴

The decision tree for this model is provided in Figure 2.

There is now *potentially* an incentive for agencies that exceed their performance target in the first period to invest in innovation, but such innovation is not guaranteed. To see this, imagine that the agency receives a bonus in the second period if second period performance exceeds first period performance ($\alpha = 1$). If the agency exceeds the performance target in year 1, then the agency received a bonus. Because of the deterministic nature of performance, an increase in budget results in an increase in performance even in the absence of innovation. Thus, agencies that exceed the threshold in the first period will exceed it in the second period and will continue to have no incentive to innovate.

However, if $\alpha > 1$, all agencies may have the incentive to invest in innovation. If the agency does not invest in innovation, there is some possibility that the agency will fail to meet the performance target in future years. If an agency does invest in innovation, certain costs are born with uncertain benefits. Thus in the second period all agencies, regardless of initial performance, must decide whether the expected benefit of innovation (in terms of the expected bonus) is greater than the costs.

This modification to the model highlights one of the strengths of the PART program. By requiring an efficiency measure and by requiring ambitious long-term targets, the PART program does provide incentives for continual investment in innovation to enhance performance. Thus, it has the potential to encourage all agencies, regardless of baseline performance, to continue to conduct evaluations and use these

⁴ This model mimics a model developed by DeGeorge, Patel, and Zeckhauser (1999) to describe managers' response to performance bonuses based on division or company earnings.

evaluations to improve future performance. Whether or not this potential manifests itself in practice will depend on the cost of innovation for each agency.

5.2. Changing Performance Metrics

Thus far the model has maintained a quite simplistic notion of performance. Agency performance has been modeled as a deterministic function of the agency's budget. There are two problems with this feature of the model. First, agency performance is likely to be stochastic. While budgets may be a key determinant of performance, some facets of performance are outside of an agency's control and thus we might expect that with the same budget and no innovations, agencies might nonetheless experience different performance levels in different years.

The second problem is that we have assumed that performance can be measured with certainty. This perfect metric can then be compared to some performance target (whether the target is fixed or changing over time). In reality "performance" is far less transparent. Much of what the GPRA and the PART programs address is establishing performance metrics.

To see how both sources of error in measuring performance affect the incentives to utilize evaluations to improve performance, the model can be modified as follows.⁵ Reported performance in each period is a function of the agency's budget in that period, B_t , and a random shock ε_t . In addition, agencies can manipulate their reported performance by adding some amount M_t (which can be positive or negative) to the performance metric. Importantly, agencies know the realization of the random shock when they choose the manipulation level but the "principal" does not.

⁵ This model follows closely from Degeorges, Patel and Zeckhauser (1999).

What do we mean by manipulation? One form of manipulation is to outright mislead. Under that definition there exists a “true” measure of performance, and agencies’ reported performance differs from that true measure by the amount M_t . But manipulation need not be intentionally deceitful. If performance is difficult to measure, agencies can manipulate performance by choosing from among various potentially valid measures of performance. Both types of manipulation are captured in this model.

Random shocks affect the performance outcome but are not costly to the agency. Manipulation of the performance metric is costly to the agency and these costs are represented by some cost function $K(M_t)$. Thus reported performance is given by: $P_t = Y_t(B_t) + \varepsilon_t + M_t$ and the agency’s budget is given by: $B_{t+1}(P_t, P_{t-1}) = B + \beta I(P_t \geq \alpha P_{t-1}) - K(M_t)$. At the end of the first period, agencies can decide to invest in innovation and the remainder of the model remains unchanged. The decision tree for this model is depicted in Figure 3.

How do these modifications affect the incentives for conducting and utilizing program evaluations? There are two main effects. First, innovation and manipulation are substitutes in this model. Performance bonuses are based on reported performance. Reported performance can be improved either by investing in innovation or by manipulation. Which method is preferred will depend entirely on their relative costs and payoffs. Both help the agent attain its own objectives (the bonus), but only innovations actually help attain the principal’s objective (improved performance).

The second effect is that for the first time, agencies may have incentives to *lower* reported performance. A few cases will illustrate this point. In the first case, imagine an agency that receives a large positive shock to performance (ε_t is large and positive), and

therefore, the agency's performance $Y_i(B_i) + \varepsilon_i$ greatly exceeds the performance target. Having such a large positive performance guarantees that the agency will receive the bonus in this period, but it also makes it very difficult for the agency to obtain the bonus in the next period if the performance target in the next period is some positive function of this year's performance. It may make sense for the agency to "reign in" performance this year by choosing a negative amount of manipulation, thereby lowering reported performance. Reported performance can be lowered to a level that still exceeds this year's target, but makes it easier on the agency to attain the target next year. The extent of manipulation in this case will depend on the costs of manipulation $K(M_i)$ and the size of the bonus.

In the second case, imagine an agency that receives a large negative shock to performance (ε_i is large and negative) and therefore the agency's performance $Y_i(B_i) + \varepsilon_i$ is far below the performance target. Positive manipulation may not be productive as the amount of manipulation required to exceed the performance target is extremely costly. In this case, it may make sense for the agency to "take the big bath" by choosing a negative amount of manipulation, thereby lowering reported performance. The agency will not be receiving the bonus this year in any case, but by lowering reported performance even further it makes it that much easier for the agency to attain the target and receive the bonus next year. The extent of manipulation in this case will again depend on the costs of manipulation $K(M_i)$ and the size of the bonus.

Only for agencies whose performance is quite near the target, does it make sense to engage in positive manipulation. For these agencies manipulation is not that costly (as the amount of manipulation is small) and manipulation ensures that the performance

bonus is received. This pattern of strategic response to bonuses in order to attain some target or threshold has been called “threshold-regarding behavior” (Degeorge, Patel, and Zeckhauser 1999). There is empirical evidence that private sector managers engage in threshold-regarding behavior with respect to salary bonuses (Degeorge, Patel, and Zeckhauser 1999). There is no reason to believe that public sector managers would not respond in a similar fashion.

5.3. Changing the Budget Bonus

The final way in which the PART program, in particular, can affect incentives for agencies to engage in performance-focused policy design is through the budget bonus. The current model gives agencies a fixed bonus for exceeding their performance targets. But, as previously demonstrated, the PART program is not that rigid in its assignment of budgets in relationship to PART assessment scores.

There are two ways in which the model could be modified to reflect this. Neither substantially changes the findings from the last iteration, so they will be discussed relatively briefly. First, the model could be modified so that agencies that exceed their performance targets have a higher probability of receiving the bonus budget. However, agencies that exceed their target may not receive the bonus and agencies that fail to exceed their targets may still receive the bonus. Recall that agencies decide to invest in innovation by comparing the expected benefit of innovation (size of the budget bonus times the probability that the innovation is successful) to the innovation cost. The effect of this change to the model is to lower the expected reward for investing in innovation for all agencies.

A second modification is to reward or punish agencies based on how far they are from the performance target. The budget for each agency is then given by:

$B_{t+1}(P_t, P^*) = B + \gamma F(P_t - \alpha P_{t-1})$, where $F()$ is a monotonically increasing function. Agencies that greatly exceed the performance target get a larger bonus and agencies that fall short of their targets receive proportionate budget cuts. The remainder of the model-- in particular, the stochastic nature of performance and the ability to manipulate the performance metric -- remains unchanged.

This modified reward/punishment scheme dampens the incentives for agencies to engage in threshold-regarding behavior. For agencies whose performance greatly exceeds the threshold, reporting higher performance now results in a greater budget bonus. This in turn makes good performance in the next period more likely. However, higher performance this year also makes attaining the performance target next year more difficult. Similarly, this modification dampens, but does not eliminate the incentive for agencies whose performance is well below the target to “take the big bath” by reporting very low performance in year one so that exceeding the performance target is easier in year two.

Whether it will make sense for the agency to engage in strategic manipulation of the performance metric will depend on the magnitude of the budget bonus multiplier γ , which determines how much better off the agency will be if it reports better performance in the current period; the variance of the stochastic component of performance ε_t , which determines the variance in performance for a given budget; and the discount rate, which determines how agencies value money today versus money tomorrow.

The key insight from these models is that none of the incentive structures examined ensures that all agencies will move toward implementing the ideal learn-act-evaluate model. The difficulties in implementing the ideal model through incentive

structures alone include problems with establishing the performance targets, problems with establishing the budget reward and punishment system, and, especially, problems with actually measuring performance. The models imply that GPRA's and PART's focus on incentive structures may move some agencies toward conducting more thorough evaluations and using these evaluations to better design programs for achieving results. But these programs may also create incentives to manipulate performance metrics. This type of activity is costly and does not result in improved government performance or improved social welfare. The concluding section will offer some suggestions for improving these initiatives.

6. Beyond GPRA and PART – Suggestions for Improving

Performance-Focused Governance

Based on this analysis, what more should be done to move agencies toward a performance-focused system? One way to reduce incentives for manipulation and increase incentives for performance-focused innovations is to absolve the agency from responsibility for shocks to performance. This requires developing performance metrics that are specific to the success of the agency's programs. For example, the annual metric for the Air Toxics program is the percentage reduction in nationwide air toxics emissions from stationary and mobile sources combined. But reductions in air toxics emissions are not entirely under the program's control. If people drive less because gas prices have risen, this may result in a decrease in mobile source emissions, but this decrease cannot be attributed to the program. This decrease may make the percent reduction easier to obtain this year, but will certainly make future percentage reductions more difficult to attain if gas prices rise and driving increases. Making agencies responsible for metrics that are not entirely (or even mostly) under their control is

unlikely to encourage them to focus on improving performance and more likely to encourage them to focus on managing the numbers.

This is not an insignificant problem for EPA programs. Of the 20 EPA programs that have already been evaluated using the PART assessment, none have metrics that are entirely within the agency's control. Most measure some level of ambient pollution level, which is subject to change in ways that are unrelated to government programs. There is little evidence from the PART reports that the goals were set in a way that could isolate the effect of the program from these other drivers of environmental change.

The PART program could be improved by focusing as much attention on measuring progress as on developing metrics of progress. While there are several questions about whether agencies have established appropriate short and long-term goals, there is only one part of one question which concerns how agencies will measure the attainment of those goals appropriately. One of the components of the question on whether the program has had an independent evaluation concerns whether the evaluation was of sufficient quality. In the guidance on this question the definition of evaluation quality focuses on randomized experiments, a method that is nearly impossible to use for many of EPA's key programs (Bennear and Coglianesi, 2004). Additional guidance and support for developing the analytic methods of program evaluation are required.

Another way to reduce the amount of manipulation of the performance metric and encourage more performance-focused innovations is to increase the relative cost of manipulation. The classic way to do this is through auditing. Independent auditors assess the validity of the reported performance measure. If manipulation is uncovered then the agency receives a penalty. Both increasing the probability of discovering or the

size of the expected penalty can deter manipulation (Becker 1968, Polinsky and Shavell 2000).

Classic auditing is best suited to control blatant misrepresentation. When manipulation takes the form of choosing the most self-serving of a set of potentially valid performance measures, third party review may prove more appropriate. Third party reviewers can provide suggestions and guidelines for how performance may be measured, thereby reducing the range of possible performance measures and reducing the potential magnitude of manipulation.

GPRA is silent on the issue of third-party review, but the PART program does require independent evaluations of each program. And the PART program itself can be viewed as a type of third-party review since OMB reviews each program's PART report and issues the final draft of the report.

The ultimate success of third-party reviews hinges critically on the independence of the reviewer (**Bazerman ???**). The concern that is most often raised about auditor or review independence is that the auditor will be too close to the agency or program that she is asked to review and will, therefore, provide too flattering a review. However, equally problematic is if the review has an agenda quite different from that of the program she is reviewing, leading to an overly negative review. In both cases, rather than have an external review help clarify what the true level of performance is, the external review may be viewed as just another estimate of performance, leaving open the question of which measure is "better."

An examination of the potential sources of external reviewers reveals that there is potential for both types of bias under the PART requirements for third-party review. There are three main sources of external reviewers for government programs. The first

is academics. But academics are motivated by what is likely to get published or of broad interest to other academics in their field. This tends to result in some programs that are heavily studied and other programs that are not studied at all. In the environmental policy realm, many more academics have examined voluntary programs and environmental management systems programs than have reviewed the efficacy of EPA's block grant programs or even its regulatory programs.

The second source of reviewers is government contractors. Contractors can be hired to conduct evaluations of all programs. However, contractor independence is jeopardized by the desire to maintain an ongoing consulting relationship with the program.

The final source of reviewers is other government agencies. OMB is currently playing the role of the third-party reviewer on all PART programs. However, OMB is not just an oversight agency. It also has an agenda, which is to help the White House develop a budget and promote the president's initiatives. This agenda may jeopardize the ability of OMB to provide objective advice to agencies on how to measure performance and how to improve performance.

One recommendation is to commission advisory groups to inform the evaluation process and serve as external reviewers. Congress established a Science Advisory Board at EPA in the late 1970s. The Science Advisory Board is composed of outside experts who provided guidance and objective review on a wide range of scientific and economic issues. Currently, there is no committee on the Science Advisory Board focused on evaluation, but this seems like a natural extension of the Science Advisory Board's role and one this is likely to provide more objective guidance on evaluation than other sources.

In addition to the promoting a system of third-party review, another recommendation focuses on building agency capacity and support for evaluations. Rather than just providing yes/no checklists, OMB and Congress can build agency capacity by offering or funding training in evaluation methods and by working cooperatively with agencies to build a social and organizational culture of evaluation. While changing an organization's culture is difficult to do from the outside, in this case working to build such a culture appears to be one of the most important characteristics of agencies that are performance-focused (GAO 2003). One concern for both GPRA and PART is that the implementation of evaluation requirements without an agency culture that supports them will result in a new set of bureaucratic procedures without substantive change in behavior. Agencies learn how to "check the evaluation box" without having to really learn from the evaluation process.

7. Conclusions

This paper outlined an ideal performance-based policy model and identified four categories of barriers to implementing this ideal model—cognitive, social, organizational and incentive-related. The paper then evaluates two programs, GPRA and PART, that were designed to reduce those barriers and move government towards a performance-focused system. We examine several different incentive structures represented in the GPRA and PART programs and evaluate how these incentive structures affect agency behavior. The results suggest that the programs cannot create incentives that will ensure that *all* agencies conduct evaluations and use these evaluations to improve performance. In particular, the fact that performance cannot be measured with certainty, but rather is subject both to external shocks and to manipulation of the performance metric, implies that agencies have a range of strategies to respond to

performance incentives. Under some circumstances, the optimal strategy will be to learn from evaluations and invest in improved performance. In other circumstances it may make more sense to manipulate the performance metric and manage to the performance target.

Auditing and third-party review can mitigate some of these perverse incentives, but the success of an auditing or review program hinges upon the independence of the auditor. This independence is questionable for at least some of the groups likely to serve as reviewers of government programs. Expanding the role of the Science Advisory Board at EPA may help provide objective third-party reviews.

In addition, more effort through training and guidance needs to be provided for agencies so that a culture of evaluation can be built. Incentives are easier to mandate than changes in organizational culture. Yet, focusing on incentives in the absence of strong social and organizational cultures of evaluation and learning risk making PART and GPRA another set of bureaucratic processes that agencies feel they must get through, rather than power vehicles for improved governance. (Also add sentence or two here about cognitive and social barriers?)

Bibliography

Anderson (2003)

Bazerman, Max H. and Michael Watkins (2004) *Predictable Surprises: The Disasters You Should Have Seen Coming and How to Prevent Them* (Boston : Harvard Business School Press).

Becker, Gary S. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76: 169-219.

Benbear, Lori S. and Cary Coglianese (2004) "Measuring Progress: Program Evaluation of Environmental Policies," *Environment*, 47(2): 22-39.

Brewer, Garry D. and Peter deLeon (1986) *The Foundations of Policy Analysis* (Homewood, Ill. : Dorsey Press).

Brown and Duguid (???)

Cohen and Sproul

Degeorge, Francois, Jayendu Patel, and Richard Zeckhauser (1999), "Earnings Management to Exceed Thresholds," *The Journal of Business* 72(1): 1-33.

Government Accountability Office (1998) "Program Evaluation: Agencies Challenged by New Demand for Information on Program Results," Report Number GAO/GGD-98-53 (Washington, D.C.: Government Accountability Office.)

Government Accountability Office (2000) "Managing For Results: EPA Faces Challenges in Developing Results-Oriented Performance Goals and Measures," Report Number GAO/RCED-00-77 (Washington, D.C.: Government Accountability Office.)

Government Accountability Office (2003) "Program Evaluation: Culture and Collaborative partnerships Help Build Agency Capacity," Report Number GAO-03-454 (Washington, D.C.: Government Accountability Office.)

Grossman, Sanford J. and Oliver D. Hart (1983) "An Analysis of the Principle-Agent Problem" *Econometrica* 51:7-45.

Hart, Oliver D. and Bengt Holmstrom (1987) *The Theory of Contracts*. In *Advances in Economic Theory, Fifth World Congress*, edited by T. Bewley. New York: Cambridge University Press.

Klayman (1995)

Kogut and Zander (1996)

Lee (1998)

Mendeloff, John (2004) "Evaluation in Two Safety Regulatory Agencies" AEI-Brookings Joint Center For Regulatory Studies, Regulatory Analysis Paper 04-05 (Washington, D.C.: AEI-Brookings Joint Center For Regulatory Studies).

Metzenbaum, Shelly (1998) "Making Measurement Matter: the Challenge and Promise of Building a Performance-Focused Environmental Protection System," Report of the Center for Public Management, Brookings Institution, CPM Report 98-2 (Washington, D.C.: Brookings Institute).

Niskanen, William A., Jr. (1994) *Bureaucracy and Public Economics* (Aldershot, U.K.: Edward Elgar).

Office of Management and Budget (2005) "Guidance for Completing the Program Assessment Rating Tool" available online at <http://www.whitehouse.gov/omb/part/>

Polinsky, A Mitchell and Stephen Shavell (2000) "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature* 38: 45-76.

Powell, Mark R. (2000) *Science at EPA: Information in the Regulatory Process* (Washington, D.C.: Resources for the Future Press).

Sprance, et al. (1991)

Tversky and Kahneman (1974)

Table 1: Ideal model of learning, action, and evaluation, and potential barriers

	Learning	Action	Evaluation
IDEAL	Gather all relevant information, including evaluations of past programs Assess validity of information Identify optimal course of action identified	Adopt optimal course of action	Collect data on outcomes Evaluate effect of program on outcomes Make evaluations available to inform future policies
BARRIERS:			
Cognitive	Reliance on heuristics (e.g., availability heuristic) Confirmation bias Status quo bias Bias toward errors of omission vs. errors of commission	Status quo bias Bias toward errors of omission vs. errors of commission	Confirmation bias
Social	Identity Sense of agency's purpose Communities of practice Cognitive dissonance: hard to accept evidence of past failures	Social rewards/punishments Identity Norms	Evaluation culture Norms that reinforce strong priors and confirmation bias Cognitive dissonance: hard to accept evidence of past failures
Organizational structure	Information sharing "Silos" Hierarchy Communication costs (time, money, effort)	Individuals with knowledge don't have authority to act Budget concerns	Budget for evaluation? Expertise for evaluation? Organizational mandate Collaborative partnerships
Incentives	Incentives to share information Incentives to search for information	Principal-agent problems: Disconnect between "optimal" action and agency's budget-maximizing choice	Incentive to conduct evaluations Fear of revealing poor performance

Table 2: Percent Change in Budget By PART Rating 2003-2005

	2003-2004		2004-2005	
	Number of Programs	Percent Change in Budget	Number of Programs	Percent Change in Budget
All programs	399	4.2%	397	-1.2%
Effective	45	13.7%	45	7.2%
Moderately Effective	105	4.7%	104	7.8%
Adequate	82	3.0%	82	-1.6%
Ineffective	19	-4.2%	19	-37.7%
Results Not Demonstrated	148	2.6%	147	-5.1%

Bold entries are significantly different from all programs in that year at the 10% level or less.

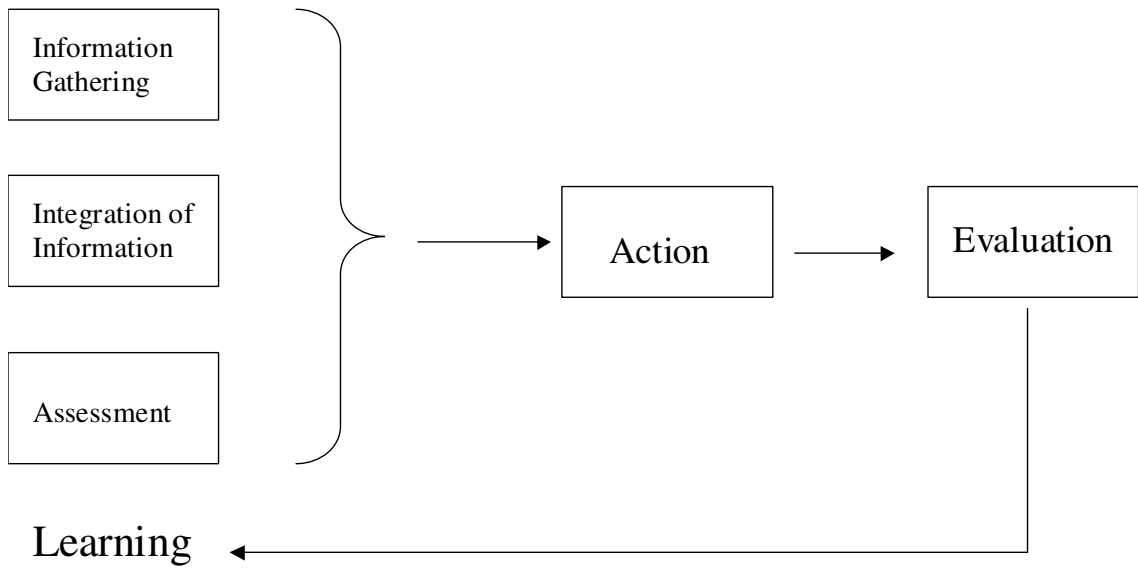


Figure 1: Ideal Model of Performance-Based Policy

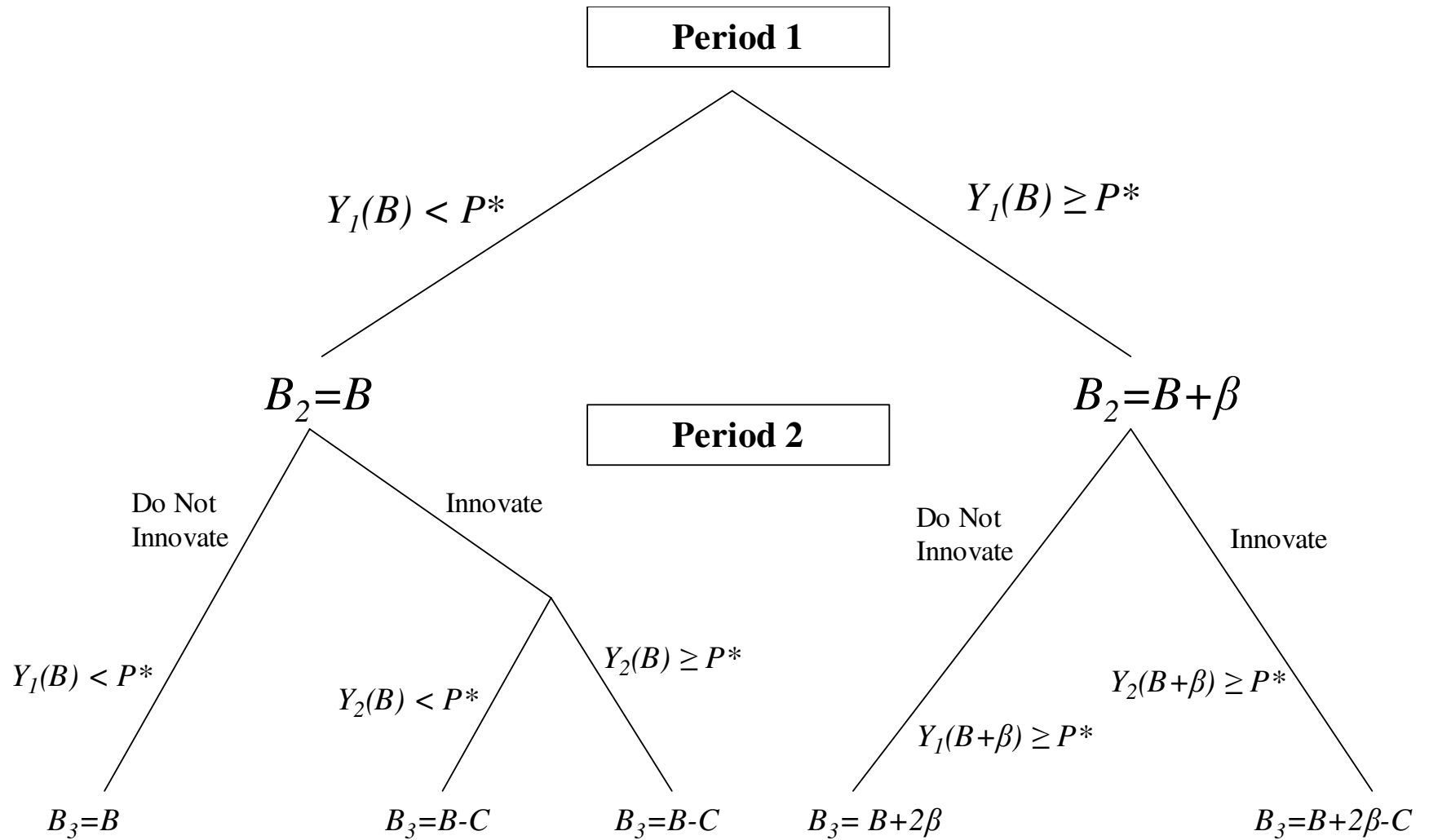


Figure 2: Decision Tree with Fixed Performance Targets and Deterministic Performance Levels

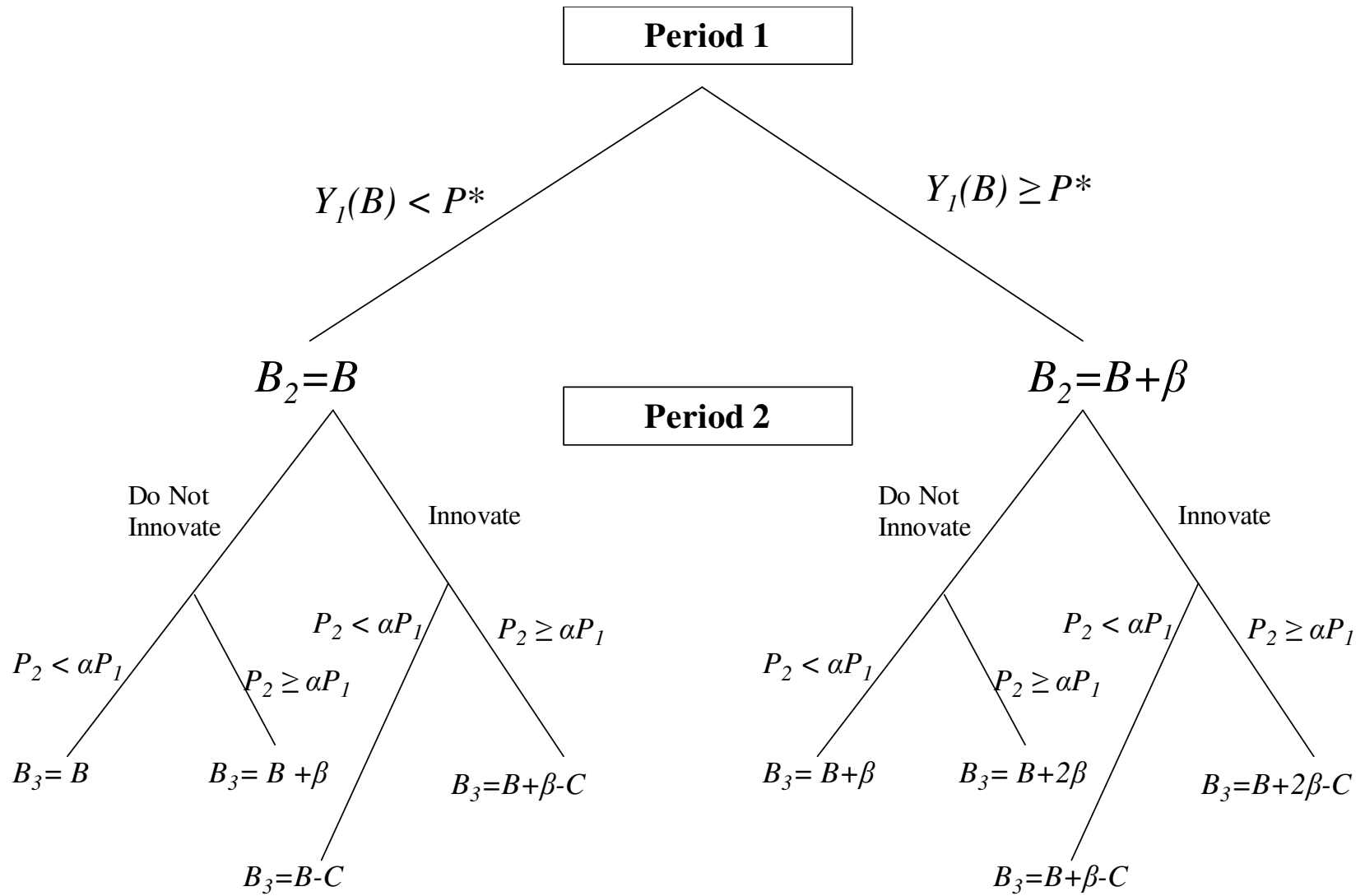


Figure 3: Decision Tree with Increasing Performance Targets and Deterministic Performance Levels

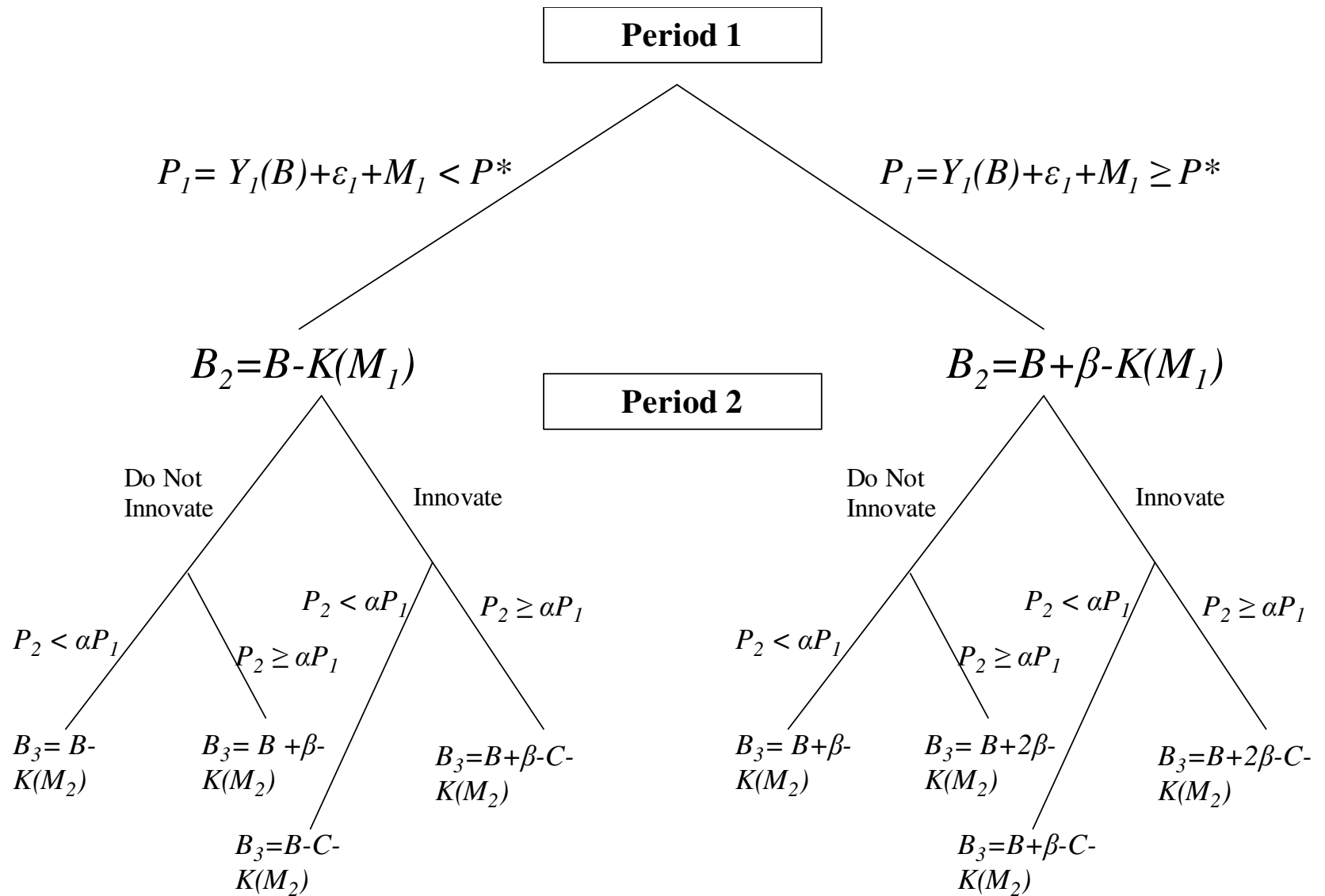


Figure 4: Decision Tree with Increasing Performance Targets and Stochastic Performance Levels