Explaining IO Performance:
An Initial Framework

Tamar Gutner
American University
tgutner@american.edu

Alexander Thompson
Ohio State University
Thompson.1191@osu.edu

Introduction[1]

International organizations (IOs) are essential but controversial actors in world politics today. They are expected to rebuild war-torn societies, reduce extreme poverty, stop the spread of disease, prevent and mitigate financial crises, address global environmental problems, adjudicate disputes, make trade more free and fair, promote gender equality, reform legal systems, and reduce corruption. These examples just skim the surface. IOs are increasingly relied upon to manage what former UN Secretary General Kofi Annan famously called "problems without passports," which states cannot easily address on their own. But instead of being praised for their contributions, IOs face relentless attacks from critics who believe they are ineffective—or worse, that they exacerbate the very problems they are supposed to ameliorate. Criticism and contradictory calls for reform are now part of daily life at most major multilaterals.

While it is widely recognized that IOs sometimes produce ineffective results or unintended consequences, the IO literature is underdeveloped in its ability to explain why this occurs—and why other IOs seem to perform quite well. Why do some IOs perform better than others? What are the determinants of IO performance? Finding answers to these questions requires us to engage in the more fundamental task of conceptualizing and measuring performance in ways that shed light on particular IOs while allowing us to generalize across them.

The question of performance has taken on additional importance in light of widespread criticism that IOs are undemocratic and therefore lack legitimacy (Dahl 1999; Held 1995; Zweifel 2006). The argument is that international institutions are too removed from individual citizens, lack transparency in their decisionmaking, and are not subject to accountability mechanisms—all features that clash with democratic norms. Given these democratic and procedural deficits, the most viable source of legitimacy for IOs in the foreseeable future is likely to be good performance. As Buchanan and Keohane (2006: 422) note, a global governance institution receives support based primarily on the extent to which it can "effectively perform the functions invoked to justify its existence." For most IOs, in other words, performance is the path to legitimacy,[2] and thus our ability to understand performance—what it is and where it comes from—is crucial.

Yet, while performance evaluation has been a hot topic inside individual organizations and in the policy literature, it has not been on the radar screen of most international relations scholars. Those seeking to expand IO theory remain focused on distinct questions of why states create institutions, how they pursue their interests through institutions, and whether and how IOs "matter." All of this scholarship appears largely removed from debates in the policy world on the performance of IOs, which tend to concentrate on more narrow issues of importance to particular IOs. Obviously, different specific criteria have to be used to evaluate individual

[2] For more discussion of performance as a source of IO legitimacy, see Cronin and Hurd 2008.

organizations; analyzing the performance of an international court is different from analyzing the performance of UN peacekeeping in the details of what is being measured. The same can be said of looking at different aspects of a single organization's performance. The UN's performance on broad issues of international peace and security must be evaluated differently from the ability of one of its agencies, the WHO, to fight the spread of malaria. Nevertheless, we believe it is possible to provide some general guidelines for thinking about performance.

In this project we seek to improve our conceptualization of international organization performance (IOP) and our ability to evaluate and explain IOP in ways that both contribute to theory and inform policy discussions. At a time when many argue that major multilateral organizations are collapsing under the weight of globalization and require rethinking and reform, it is critical that scholars do more than sit passively on the sidelines.

This paper proposes an initial framework for understanding important aspects of IOP. We examine how it has been addressed (mostly indirectly) in the IO literature, offer a conceptualization of performance as applied to IOs, and suggest several analytical tools as a starting point for discussion and research. We argue that "performance" can be understood as both outcomes and process, and that this distinction also helps us think about different ways to measure performance. For some organizations, we are interested in measuring their effects on the real world (outcomes) whereas for others we should focus more on process indicators and intermediate outputs as reflections of performance. There are tradeoffs associated with these conceptualization and measurement choices; neither is categorically superior. We also provide a roadmap for thinking about the sources of good and bad performance, which can be internal to the organization (such as bureaucratic dysfunction), external (such as the role of power politics), or both. Finally, we suggest research strategies for studying IO performance, addressing the tricky issues of how to establish baselines against which to assess performance and how to disentangle the causal role of the IO from other exogenous factors.

Performance in the IO Literature

Despite increased attention, both theoretical and empirical, to the role of formal intergovernmental organizations in world politics (Abbott and Snidal 1998; Gartzke, Hafner-Burton, and von Stein 2008), few IO scholars have looked directly at the issue of performance. A partial exception is Barnett and Finnemore's (1999, 2004) pathbreaking work, which identifies bureaucratic dysfunction as a key source of undesirable—even "pathological"—behavior. Building on Weber's arguments about domestic bureaucracy, they argue that IOs are "social creatures" that use their authority, knowledge and rules to act autonomously in ways that may or may not reflect the interests and mandates of states. For example, "bureaucracies may become obsessed with their own rules at the expense of their primary missions in ways that produce inefficient and self-defeating outcomes" (Barnett and Finnemore 2004: 3)**.**

Bureaucratic dysfunction is clearly an important source of poor performance. However, there are limitations to Barnett and Finnemore's framework when it comes to understanding IOP. First, not all IOs have substantial bureaucracies. One study of regional economic organizations finds that some do not even have meaningful secretariats; for those that do, the staff possess very little in the way of resources and discretion (Haftel and Thompson 2006). Thus to the extent that we want to move beyond the "usual suspects"—the UN, IMF, World Bank, EU, etc.—we need a broader set of considerations. Second, in a related way, Barnett and Finnemore do not go very far in explaining *variation* across IOs and its implications. Why do some IOs exhibit

3

pathological behavior and poor performance while others do not? Presumably there are meaningful and systematic differences across IOs that help explain these different outcomes. Finally, their work offers a micro-perspective on sources of disconnect between broader norms and interests and an IO's entrenched bureaucracy. But they downplay the *external* side of the equation—the impact of power politics and state interests on the organization and its ability to carry out its tasks.

The rationalist literature on IO design and delegation provides this focus on variation and external factors by explaining institutional outcomes as a function of underlying cooperation problems (Koremenos, Lipson, and Snidal 2004; Hawkins et al. 2006; Pollack 2003). Institutions in this view are the product of state interests, and the ability of IOs to act independently to shape outcomes is a function of their relationship with states, especially the nature of the initial delegation "contract" and of the control mechanisms established by states. However, for the most part this literature implicitly assumes that *efficient* designs are also *effective* designs. They do not take the next step to ask the all-important question of whether a given design leads to better performance and improved outcomes (Wendt 2001). We hope to incorporate insights from the design and delegation literature but to treat these characteristics of IOs as independent variables that help explain IOP rather than as dependent variables.

Finally, the vast literature on "regime effectiveness" also offers insights that are helpful for the study of IOP. This literature drops the rationalist-institutionalist assumption that theoretically efficient designs are necessarily effective, explores in more detail the concepts of effectiveness and outcomes, and recognizes and seeks to explain their variation (Young 1999; Underdal and Young 2004). While the literature includes a number of definitions of effectiveness, a common one is whether the regime "solves the problem that motivated its establishment" (Underdal 2002: 11). For example, for environmental regimes the question is whether they improve the physical state of the environment. The degree of success can be measured in two ways: How much improvement do we see compared to a scenario with no regime? And how close is the outcome to the optimal solution to the problem (Helm and Sprinz 2000)?

While the regime effectiveness literature offers suggestions on defining and measuring institutional effects, its application to IOP has limits. First, a given IO is only one component of a regime, which may include multiple IOs and other institutions—including norms, treaties and non-governmental actors—that share in the governance of a given issue-area. Thus only rarely can outcomes in a given regime be attributed to the performance of a single IO. This is clear in the context of international standards regimes, which are multifaceted and governed by a mix of private, governmental and intergovernmental actors (Abbott and Snidal, this project). Second, most work on regime effectiveness has focused on environmental issues and so we must be cautious about whether the lessons travel well to other issue areas. In particular, there are reasons to believe that the tangible and observable metrics (number of whales, emission of pollutants, etc.) in that domain may not exist in others. Most organizations pursue goals that are amorphous and provide more intangible services, making "surrogate quantitative measures of organizational performance" hard to come by (Forbes 1998: 184). While all three of these literatures—on IO bureaucracies, rational delegation and design, and regime effectiveness—are helpful for understanding IOP, none addresses it directly. An important gap in the literature remains.

The Concept of Performance

When we discuss IOs, we are referring generally to intergovernmental organizations that are comprised of two or more member states, established by agreement among governments or their representatives, and sufficiently institutionalized to involve regular meetings, rules governing decisionmaking and a permanent staff.[3] Performance in this context thus refers to the behavior of the relevant actors (the member states and staff) and to the effects of the institutional structures of the organization itself. This gives us a sense of who and what are doing the performing, with particular actors and structures taking on increased relevance in a given case. The *causes* of good or bad performance may come from within the IO or from external sources, a distinction we develop below.

Performance as an explanandum is a multifaceted concept. In everyday usage it has two distinct but related meanings. First, as a verb, to perform is simply to fulfil an obligation or complete a task. Second, as a noun, performance refers to the manner in which a task in completed. Thus to address the issue of performance, as applied to the social world, is to address both the outcomes produced and the process—the effort, efficiency and skill—by which goals are pursued by an individual or organization. Conceptually and empirically, these two senses of performance are closely related: We should expect highly capable and efficient individuals and organizations to complete tasks and attain goals more effectively.

We apply both the outcomes and process aspects of performance to the study of IOP. This mirrors conceptualization of organizational performance in the business management, public administration and organizational theory literatures, where it is most developed. Thus, a simple starting point for defining performance is that it refers to an organization's ability to achieve agreed-upon objectives. This is a useful way to think about performance when objectives are fairly well defined, as in the case of corporations (with their clear objective to achieve profits), but is less straightforward when goals are more ambiguous and variegated. This is true for almost all public organizations, both governmental and nonprofit. In such cases, there may be different definitions of what constitutes goal achievement, reflecting the attitudes of various participants and observers toward the organization's results and even underlying disagreement over what constitutes a good outcome (Kay and Jacobson 1983). We might therefore be as interested in performance as captured in internal processes, including the ability of the organization to mobilize resources and make internal operations more efficient (Kaplan 2001; Simon, Smithburg, and Thompson 1973).

Even if goals are well-defined, however, it should be noted that well-functioning internal processes do not necessarily imply that an entity will fulfil its goals. The expectation may simply be too great or the organization's goals too difficult to achieve. By the same token, if goals are easy to achieve an entity might succeed perfectly well even when its performance per se is not very impressive. This helps us understand why *performance* is distinct from *effectiveness*, because the latter implies an ability to achieve specific outcomes or to solve problems without reference to the underlying capacity of the entity, the impact of complicating constraints, or the manner by which outcomes are achieved. Performance is more contingent and variegated. Also, in practice, most organizations and outside observers judge performance at numerous levels, setting goals and devising indicators for specific tasks that together offer a holistic measure.

---

[3] For a similar definition, see Shanks, Jacobson and Kaplan (1996). While we limit our discussion to formal IOs, our framework might apply usefully—albeit in a less straightforward manner—to the performance of international NGOs or more informal intergovernmental institutions. We leave this possibility to future research.

5

The Challenges of IO Performance Analysis

IOs share many characteristics with other types of complex public organizations and thus face some of the same challenges when it comes to defining performance. Most importantly, they are trying to achieve multiple and sometimes conflicting goals and thus are being pulled in different directions by stakeholders with various degrees of power and influence (Perrow 1986: chapter 7; Moe 1989; Markus and Pfeffer 1983). IOs also share with other public organizations the reality that many of these goals are political, broad or ambiguous in nature, and by definition the achievement of these goals is difficult to measure objectively.

As a result, in the real world, outside neat conceptual boxes, defining performance for IOs is especially messy and political. First, in addition to serving multiple functions, IOs commonly have lofty mandates that do not offer specific criteria for judging performance. When examining a company's performance, we begin with its financial statement. How does this compare to an institution like the World Bank, whose major goal is to "work for a world free of poverty"? Often IO objectives are broken down into multiple specific categories, and scholars and practitioners may highlight a subset and present it as evidence of overall good or poor performance. This is particularly evident in major IOs. For example, the World Bank's performance has been judged by various internal and external observers by looking at its overall financial performance as a financial institution; the performance of individual projects or programs or sectors; the existence and quality of a wide range of specific policies, strategies, and due diligence procedures; or evidence that those policies, strategies, and procedures are being followed and have a lasting impact. The agreed-upon objectives are often numerous or ambiguous—or both. Discussion of the "UN's" performance is almost nonsensical; it may be impossible to come up with an aggregate metric of the performance of a body that has so many disparate parts and goals.

Second, given the complexity of their often expanding tasks and the number of principals and other constituencies they must please, IOs are unusually prone to what we call the "eye of the beholder" problem, in that analyses of their performance vary significantly depending on the analyst. Bankers may be happy with the IMF's bottom line results, while NGOs howl at the institution's inattention to poverty reduction or environmental standards. Does the fact that the United States went to war in Iraq in 2003 without the Security Council's blessing reveal an irrelevant or a properly functioning Security Council? Both arguments have been made (Thompson 2006). This problem does not simply reflect different perceptions among "outsiders" to an IO. For example, members of the International Whaling Commission have opposite perceptions of its performance, depending on whether they are whaling or non-whaling nations. The bottom line for member states is ultimately whether or not they perceive that they benefit from the IO both individually and in terms of collective gains, and that these benefits could not be achieved through some alternative arrangement. Individual member states may also distinguish between an IO's broad official goals and its operative goals (what the organization is really trying to do), and not mind if one category is not being met, as long as the other is (Perrow 1961; Wilkinson, this project); Lipson (this project) points out that in some cases a poorly-performing institution may in fact be *desirable* for key stakeholders and conducive to organizational survival. The fact that there are violently opposed perceptions on the performance of virtually any major IO makes it even more important for scholars to offer better ways of conceptualizing performance.

Third, and related, is the fact that one of the "beholders" involved in performance evaluations is, obviously, the IO itself. Virtually all IOs evaluate themselves, offering performance objectives and ways of measuring them, and internal evaluation mechanisms are increasingly common, especially in large multilateral organizations. While it is tempting to rely on such evaluations—and, indeed, they might be helpful as a first cut—it also poses potential problems. To begin with, an organization's bureaucrats have their own interests and biases, which may prevent objective evaluations and may even lead to self-serving ones, designed to justify past decisions and to cast internal actors in an attractive light. Beyond these narrow interests, bureaucrats understand that their organization faces competition from others, and this creates an incentive to provide overly positive evaluations to stakeholders, funders and political principals (Powell 1991; Cooley and Ron 2002). Finally, internal evaluation, especially in the context of budget constraints, sometimes involves shortcuts of convenience. Managers naturally prefer "to measure aspects of their programs that are amenable to measurement" rather than to devise more complicated—but potentially more accurate—indicators of performance (Kelley 2003: 857). Precisely because most IOs are serving multiple goals and stakeholders, internal evaluators have a variety of measures at their disposal, which can be relied upon and manipulated in line with these incentives.
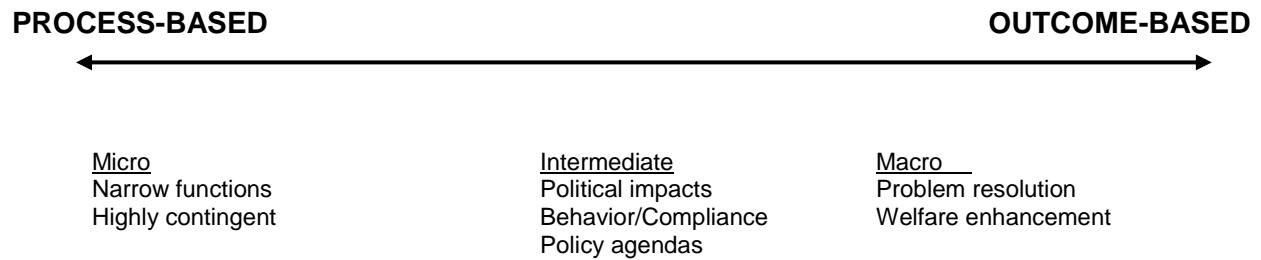
The issue of IO self-evaluation is one that international relations scholars are only just beginning to study. Weaver (this project) points out the "paradox" of self-evaluation, in that independent evaluation units within IOs are supposed to both "steepen the organizational learning curve" but also "enhance external credibility or legitimacy." These may be irreconcilable. Lipson (this project) reminds us that organizational goals themselves are inherently political, since they are often the result of bargaining and negotiation among different "organizational constituencies." Conceptualizing "good" performance in this context is at best political and at worst completely arbitrary and counterproductive.

While many of the obstacles to performance analysis are ubiquitous and common to organizations of all types, IOs are unusual in that they are governed by the world's states, and hence one important aspect of their behaviour and performance can be traced to the ability of governments to cooperate internationally and effectively manage large organizations. This is made more complicated by the fact that states themselves have an awkward and fundamentally liminal relationship to IOs, with their representatives (ambassadors, delegates, etc.) situated inside the organization and their capitals sitting outside, with avenues of communication and influence in both directions (see Elsig, this project). Governments also face pressures at home, and thus may not be consistent in their interactions with and within IOs. As we conceptualize IO performance and explore its causes, we are constantly reminded of the unusual and relatively anarchic setting offered by international politics. Arguably, understanding and explaining the performance of *international* organizations is uniquely difficult—and uniquely interesting.

Metrics of Performance

To help organize the various possibilities, we offer a continuum of metrics for evaluating IOP, with macro-outcomes at one end and more process-based indicators at the other (see Figure 1). This offers the possibility of considering a wide variety of measures rather than imposing a single metric. As an initial move, this facilitates one goal of the project, which is to determine what approach or mixture of approaches is most promising and under what circumstances.

Figure 1. Potential Performance Metrics

PROCESS-BASED   **OUTCOME-BASED**

<--->

Micro   Intermediate   Macro
Narrow functions   Political impacts   Problem resolution
Highly contingent   Behavior/Compliance   Welfare enhancement
                               Policy agendas

One possibility, at the right end of the continuum, is to look at macro *outcomes*. As we have noted, this is the approach adopted in much of the literature on regime effectiveness, where effectiveness is often defined in terms of problem solving and measured by aggregate outcomes. The outcome metric is arguably the most intuitive way to evaluate institutional effects. If outcomes can be measured, this approach holds the promise of allowing for comparison across IOs and hence for generalization.

However, as we noted above, performance measured in terms of outcomes may not be appropriate when IOs are constrained by various political and other factors outside of their control; in such cases, it is difficult to link outcomes causally to the role of the IO. For example, the UN High Commissioner for Human Rights may work hard to offer technical support, training, norm diffusion, and so on, in hopes of encouraging or shaming states to halt abuses, but the UNHCR itself cannot be held accountable for cases where abuses continue despite its efforts. Also, outcome metrics apply most easily to organizations that have a single, clear objective. For these reasons, we expect that outcome-based metrics are most appropriate in circumstances where a fairly autonomous IO plays a predominant role in a given regime, and in issue areas with objectively definable and measurable solutions.

At the other end of the continuum, by contrast, we might analyze the *process* of IO behavior and decisions at a micro level. Here we focus on the specific tasks and narrow functions the organization is intended to perform and assess whether these are successfully carried out. Though we should not blame the Northwest Atlantic Fisheries Organization for stock depletion, we can ask whether it efficiently collects data on catches and monitors the fishing activity in its jurisdiction. This approach allows us to appreciate the context of IO behavior and observe contingencies that constrain its conduct, though our conclusions are likely to be more limited to a given IO studied in depth. It also may reveal interesting cases where IOs properly fulfill their various tasks and functions but have little or no impact in addressing the fundamental problem behind their creation.

A final possibility is to split the difference between process and outcomes in order to focus on intermediate products of IO activities. For example, some authors have suggested the use of "observable political effects" of institutions, short of aggregate impact, to assess effectiveness (Haas, Keohane, and Levy 1993). The public policy literature refers to these as the

"outputs" or "intermediate outcomes" of government programs (Levy, Meltsner, and Wildavsky 1974 ; NAPA 2002: 12).[4] For example, we might measure the number of states and localities that have implemented environmental programs required by federal legislation. At the international level such political impacts include state compliance, policy change, and the emergence of ideas and behaviors consistent with institutional goals (see also Young 1999). While these effects do not always lead to problem solving, they are likely to be associated with organizations that perform well and are often the subject of study in their own right, as in research on compliance with international institutions (e.g., Simmons 2000; Mitchell 1994).

Figure 2 helps us understand these analytical tradeoffs by portraying the various stages of IO performance as a pyramid. This illustrates places at which performance may be observed and assessed, rather than offering one overarching set of independent or dependent variables. (In fact, it may be that different sets of variables operate at different levels of the pyramid.) The stages allow us more analytical traction, because we can identify specific areas where performance is amiss and can better understand how one stage of performance impacts another. In general, we expect good performance to "trickle up," with success at each lower stage serving as building blocks for success as we move up the pyramid. At the bottom of the pyramid are the many specific tasks, projects, and programs performed by an IO. Achievement of these functions should lead to better performance at the next level, where a smaller number of intermediate outputs and political goals are achieved. For example, if the WTO secretariat produces credible Trade Policy Reviews this should promote increased compliance with free trade norms. Moving to the top of the pyramid, if administrative and political tasks are performed well, this should produce outcomes that solve underlying problems and enhance welfare.

We might therefore distinguish between *process performance* and *outcome performance*, each worthy of study in particular situations. The limitation of focusing on process performance is that it does not necessarily translate into outcome performance. This might occur either because the IO's operations are not sufficient or well suited for solving underlying problems, or because its administrative successes are offset by intervening variables at later stages, such as political clashes among states. As one study finds, improved management techniques and increased efficiency within a government agency might not lead to better outcomes for citizens (Kirlin 2001). At best, process performance is a necessary but not sufficient condition for favorable outcomes. As two public administration scholars note in reference to output (as opposed to outcome) measures, "The difficulty is that the more convenient measures do not necessarily evaluate the most important consequences of policies and/or programs" (Nakamura and Smallwood 1980: 77). Therefore we might understand how an IO performs its various administrative tasks but still not know whether it is ultimately effective in achieving its goals. Indeed, process can also be used by IOs to mask substantive outcomes, as we noted above in our discussion of self-evaluation.

The main limitation of focusing on outcome performance is that it says little about causation: We cannot know if problem solving is a function of efficient and skillful behavior on the part of the IO (its staff or member-states) or of sound institutional design. Good outcomes may not be attributable to IO-based variables. By the same token, poor outcomes may occur despite a very high level of performance, at least at certain stages. Put more generally, studying outcomes alone does not allow us to evaluate the contingent and relative nature of performance.

---

[4] For more discussion of this terminology, though with somewhat different categories, see Underdal 2002; and Easton 1965: 351-2.

Figure 2. Pyramid of Performance: From Process to Outcomes
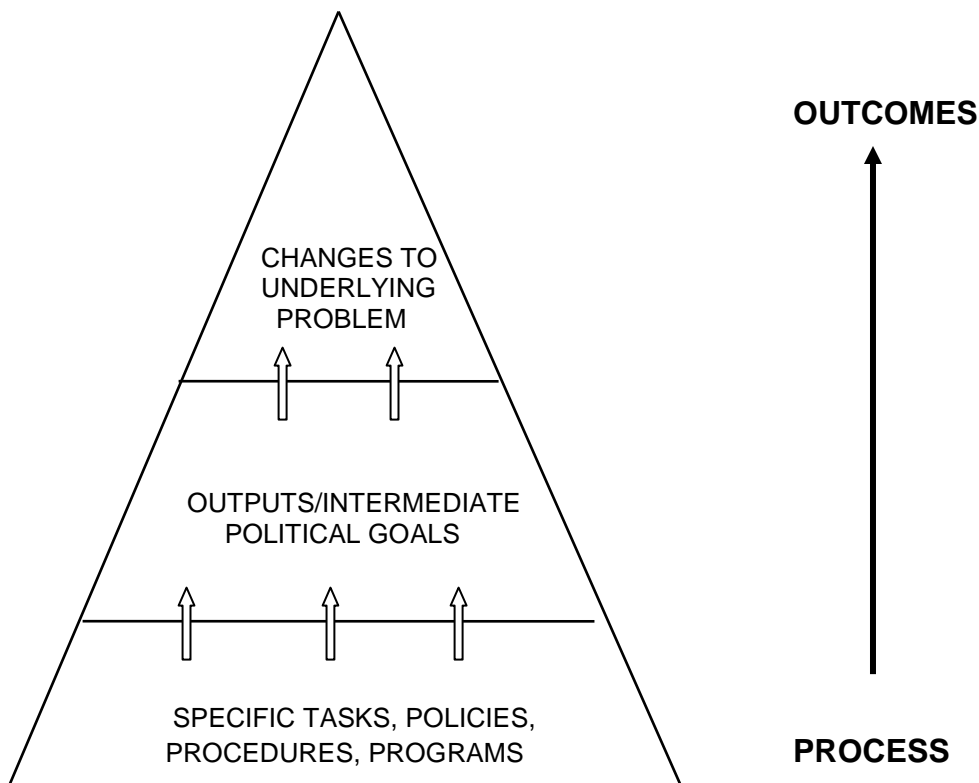
The Sources of Performance

A central question for scholars interested in IOP involves the determinants of good and bad performance. Drawing again on the existing IR literature, some of which was reviewed above, we can discern two broad approaches to thinking about the sources of performance. Beginning with the Neoliberal tradition and extending through work on design and delegation, some view IOs as subject to the design decisions and control of states (Keohane 1984; Koremenos et al. 2004; Pollack 2003; Hawkins et al. 2006). These largely rationalist approaches are functionalist in the sense that IOs are "structures of voluntary cooperation" that produce mutual benefits by helping member states to overcome cooperation problems (Moe 2005: 215)**.** IOs in this tradition are member-driven. Although they may have some autonomy, independent behavior is either consciously intended by their state principals or carefully constrained (Abbott and Snidal 1998; Garrett, Kelemen, and Schulz 1998).

A variation on this perspective begins with the assumption that IOs are mainly controlled by states, but instead emphasizes the undesirable and inefficient outcomes that may occur when IOs struggle to cope with incoherent mandates, the irreconcilable political demands of member states, and state behavior that undermines their ability to perform. Secretariats are ultimately dependent on the funding and political support of their member states, and this means that poor

outcomes may not be the fault of IO agents but the result of "principal problems" emanating from member states (Thompson 2008). For example, Vreeland (2006) argues that some of the IMF's weak performance can be attributed to pressure by its powerful shareholders to make loans without strictly enforcing the policy conditionality attached to them. Gutner (2005) suggests that one reason for the gap between the World Bank's environmental policies and its weak efforts to improve its environmental performance is that its member states have delegated conflicting and complex tasks that are difficult for the institution to institutionalize and to implement. More generally, the "green" performance of development banks is largely a function of the level of environmental commitment among its large donor states—a matter of political will largely outside of the bureaucracy's control (Gutner 2002). In sum, this first approach views IO outcomes, whether good or bad, as rooted in external, material forces.

By contrast, the second dominant approach looks within IOs to find the factors that either enhance or (usually) detract from outcomes. Barnett and Finnemore's (2004) work on the dysfunctional behavior of IO bureaucracies is the best example. In his study on the UN's role in Rwanda, Barnett (2002) points to aspects of the UN's internal culture that generated a "collective mentality" of denial that genocide was occurring; viewing the conflict as merely a civil war fit more comfortably with standard procedures based on impartiality and consent. Leadership characteristics are also an important variable within IOs, as the Paul Wolfowitz' effect on staff morale at the World Bank illustrates. The United States accused Hans Blix of being too "diplomatic" in his relations with Iraq and thus of not fulfilling the potential of the pre-war weapons inspections regime. IO behavior in this tradition is mostly a function of internal and social forces.

The evolution of the literature suggests a dichotomy when it comes to the determinants of IO performance: external-material versus internal-social. However, it is clear that other possibilities exist. While most of the internal dynamics identified in the literature are ideational or cultural, strategic calculations and self-interest also play a role within IOs. IO staff may pursue carefully calibrated strategies in order to achieve their distinct goals (Alter 1998; Vaubel 2006; Hawkins and Jacoby 2006). Often these goals are driven by bureaucratic self-interest and involve some sort of tangible, material gain, such as expanded discretion, new resources, or career success. Applications of principal-agent theory to IOs typically assume that if institutions are not achieving the desired policy outcomes delegated by state principals, it is because the agents are pursuing self-interested behavior that deviates from expectations (Martens 2002). IO staff and member-states may find themselves working at cross-purposes as a result, with the IO agent as the culprit. In some cases, the staff may simply lack adequate personnel and resources to perform well, a condition shared by many IOs.

By the same token, external influences need not be material or formalized. Borrowing from modern organization theory and sociological institutionalism, Finnemore (1996) emphasizes the possibility that IOs are a product of their social and cultural environment rather than functional efficiency. In this spirit, Roland Paris (2003) argues that a global shift toward liberal values led to the increased willingness to intervene via multilateral peacekeeping in the 1990s. In some cases, IOs may perform poorly because their missions do not reflect a clear consensus among states of what underlying normative principles should be pursued. Similarly, member-states may lack a common understanding of the underlying problem or how it should be solved. For example, efforts by the UN to tackle human rights have been plagued by different views on human rights norms and on the fundamental question of whether the notion of "universal human rights" even exists (Mingst and Warkentin 1996). On a more routine level,

IOs are constantly subject to the difficulties of cross-cultural negotiations and national pride (Cohen 1991). All of these situations risk leading to low levels of support and counterproductive activities on the part of states.

External problems may also encompass the context in which an IO operates, on-the-ground, in trying to carry out its work. An IO may be well designed, have a clear mandate, strong support from its member states, and possess a clear set of processes and procedures for carrying out its work, but still fail if it is working in a domestic situation where there is instability, weak capacity, corruption, or a lack of consent from relevant domestic parties (Gutner 2002; Howard 2008).

Much of the literature separates these variables and privileges one set over another, although more scholars are looking carefully at how these variables interact with one another and at different stages of an IO's policy and performance processes. This discussion suggests four possibilities for thinking about the sources of poor IOP, summarized in Table. 1. Barnett and Finnemore (2004, 36) offer a similar typology.

Table 1. The Sources of Poor Performance

	Internal	**External**
Social	Examples: -Organizational culture -Bureaucratic dysfunction -Leadership deficit	-Competing norms -Lack of consensus on problem
Material	-Inadequate staffing, resources -Bureaucratic/career self-interest	-Power politics -Incoherent mandates -Short-sighted principals -"On-the-ground" constraints

While it is analytically convenient to separate them, these forces obviously overlap in the real world. For example, while the effects of leadership might be an internal matter, the selection of leaders and the degree of accountability to which they are held is largely an external matter driven by member-states (Kahler 2001; Johns 2007). Agency theory offers an analytical tool to examine how both external power politics and internal bureaucratic incentives to impact IO behavior. Instead of simply focusing on states and their interests, it is a fundamentally relational theory, in assuming that principals (usually member states) and agents (usually the IO) exist in a mutually-dependent relationship with one another (Reinalda and Verbeek 2004).

Others argue for more attention to ways in which traditional principal-agent theory and sociological organization theory may complement one another to more powerfully explain how external and internal influences impact an organization's behavior. In Kate Weaver's view, P-A theory emphasizes the dominance of external factors shaping an organization's policies and operations, while sociological theory is better suited for investigating how bureaucratic culture

and politics influence an IO's practices (Weaver 2007). Similarly, Michael Lipson (2007) uses the concept of "organized hypocrisy" to describe the reaction of UN bureaucracies to often conflicting outside pressures; the resulting outcomes, sometimes positive and sometimes negative, flow from a combination of external and internal dynamics.

Finally, it should be noted that despite differences in emphasis and conceptual approach among the scholars cited above, they are united in treating IOs as important actors and in taking their details and behaviour seriously. We pick up these debates to theorize performance in a manner that is both conceptually coherent and useful for policy. We expect to see a combination of factors—social and material, internal and external—driving outcomes.

One way to summarize this discussion is that external and internal factors and social and material factors should not be assumed to be dominant or separate. How they operate together to influence an IO's performance will likely vary depending on the IO and issue. Stated differently, we suggest that the external/internal and social/material categories are candidate independent variables to be evaluated in determining an organization's performance. Performance in some cases may be a story of power politics; whether and how the major powers agree to and pursue an issue will shape performance (Wilkinson, this project). There may also be cases where member states' interests are aligned, yet something happens inside the bureaucracy to throw performance off-track, whether it be the challenge of juggling too many mandates or a mismatch between member-state goals and the incentives of bureaucrats (Hafner-Burton and Pollack, this project). Finally, as noted above, there may be cases where the political and institutional factors are aligned, but the IO is working in the midst of a difficult domestic context, such as civil war, where all good intentions are thwarted by conditions of destruction.

IOP Research Strategies

Regardless of what metric is used and what sources of IOP are being investigated, the key issue analysts must confront is how to frame what it is they are evaluating. One reason we see wildly different analyses of the same organizations is that the metrics of performance, the time period, and the tasks or objectives under scrutiny differ. We cannot resolve these problems but we can make them more transparent and render research findings more comparable across cases.

To this end, we propose the following guidelines, organized roughly in stages of research, for those studying IOP:

- Establish a baseline for assessing performance. This may be in reference to an IO's original mission, which reflects what states intended when they create the IO. Given that most IO missions expand over time, a baseline may also refer to an IO's mission at a specific point in time. Scholars can narrow ambiguous or congested missions and address the "eye of the beholder" problem by selecting specific objectives or considering performance from the perspective of a key constituency. Establishing a baseline is important because it is only against a particular set of objectives and in the context of a given timeframe that performance can be assessed.

- Specify what indicators will be used to assess performance. The researcher should explicitly link these indicators to the baseline identified in first stage by explaining how they capture performance results of interest. One place to begin is how the IO in question assesses itself. To the extent that such efforts are systematic they might provide a useful

starting point for finding relevant indicators, even if the conclusions drawn from them cannot be taken at face value for all of the reasons noted above.

- Be clear about the level or levels of analysis that will be examined, and justify this choice. The distinction between process performance and outcome performance provides a starting point, as does the three stages identified in the performance pyramid: narrow administrative functions, intermediate political goals such as compliance, and broad outcomes that contribute directly to problem-solving. In all cases, however, we should be cognizant that these various levels interact with each other and rarely tell the whole story of performance.

- Identify and analyze the sources of good or bad performance. A first step is to identify these sources as external or external, and then to discern the specific factors and mechanisms that shape IOP. In many cases, we should expect to see a combination of factors—social and material, internal and external—driving outcomes. Based on these findings, the analyst might usefully reflect on the implications for IR theory and the IO literature.

Two more general problems confront all efforts to explain IO performance empirically and do so at every stage of analysis outlined above. First, in most cases we have to take into account the difficulty of the underlying problem. Some problems are simply more complicated than others, for political or technical reasons. These differences must be controlled for when comparing performance across cases and especially across issue areas. Second, assessing performance suggests that we consider and attempt to answer an important counterfactual: What would the outcome have been absent the IO or with a different institutional arrangement?[5] While this hurdle is less relevant if we use the narrow functions of IOs as our metric of performance, in the case of intermediate political impacts and macro outcomes addressing this counterfactual is key. In some cases, the researcher can take advantage of a natural experiment or even devise an experiment to make counterfactual claims (Hyde, this project). When this approach is impractical, counterfactual analysis is likely to require process tracing to link the activities of IOs with the relevant outcome. Finally, longitudinal studies that analyze outcomes before and after an IO is created or involved might be helpful.

Conclusion

We have not sought to provide a theory of IOP. Instead, we offer suggestions on now to conceptualize, measure, and evaluate IO performance. Developing these concepts and metrics is an important first step in a new research program with relevance to both theory and policy, and we hope it will help to organize and guide further theorizing and empirical research on the subject.

Our framework raises a number of questions that can guide empirical research and further theoretical refinement, including the generation of more specific hypotheses. First, is performance best conceived of and measured with process-based or outcome-based indicators? What are the trade-offs in particular empirical contexts? Second, what are the sources of good

---

[5] Bernauer (1995: 360) makes a similar point in the context of regime effectiveness. Some of the contributions to Haas, Keohane and Levy (1993) are guided by similar counterfactual strategies.

and bad performance for IOs? Are they internal or external, social or material, or a combination thereof? Third, what obstacles do IOs face when it comes to evaluating themselves effectively? In all of these cases, we should consider scope conditions; that is, we should ask ourselves whether our conclusions are broad and generalizable or whether they apply to a more narrow set of IOs or issues.

Our workshop group on IO performance has brought together scholars seeking to address these issues through a variety of theoretical perspectives, international organizations, and issue areas. Together, we hope to develop better analytical tools and theories on the determinants of IO performance and the conditions under which given factors are likely to matter most. We hope our collaboration will generate a set of empirical studies that not only contribute to a theory-building process, but also have policy-relevant conclusions. We want to narrow or close the gap between the academic literature that ignores IO performance, and the policy literature and debate that address it in ways that are more normative than analytical.

Performance has been almost entirely overlooked in the IO literature, even though the importance of IOs in solving global problems—and thus their value to the international community—is largely a function of their performance. If they do indeed play an important role, if they perform well, this might serve as a central basis of legitimacy for IOs, which lack many of the intrinsic characteristics on which legitimacy is often based in political institutions.

References

Abbott, Kenneth W., and Duncan Snidal. 1998. Why States Act through Formal International Organizations. *Journal of Conflict Resolution* 42 (1):3-32.

Alter, Karen J. 1998. Who Are the "Masters of the Treaty"? European Governments and the European Court of Justice. *International Organization* 52 (1):121-147.

Barnett, Michael N. 2002. *Eyewitness to a Genocide: the United Nations and Rwanda*. Ithaca: Cornell University Press.

Barnett, Michael N., and Martha Finnemore. 1999. The Politics, Power, and Pathologies of International Organizations. *International Organization* 53 (4):699-732.

———. 2004. *Rules for the World: International Organizations in Global Politics*. Ithaca, N.Y.: Cornell University Press.

Buchanan, Allen, and Robert O. Keohane. 2006. The Legitimacy of Global Governance Institutions. *Ethics and International Affairs* 20 (4):405-37.

Cohen, Raymond. 1991. *Negotiating Across Cultures: Communication Obstacles in International Diplomacy*. Washington, D.C.: United States Institute of Peace.

Cooley, Alexander, and James Ron. 2002. The NGO Scramble: Organizational Insecurity and the Political Economy of Transnational Aid. *International Security* 27 (1):5-39.

Cronin, Bruce, and Ian Hurd. 2008. *The UN Security Council and the Politics of International Authority*. New York: Routledge.

Dahl, Robert A. 1999. Can International Organizations Be Democratic? A Skeptic's View. In *Democracy's Edges*, edited by I. Shapiro and C. Hacker-Cordón. New York: Cambridge University Press.

Easton, David. 1965. *A Systems Analysis of Political Life*. New York,: Wiley.

Finnemore, Martha. 1996. *National Interests in International Society*. Ithaca, N.Y.: Cornell University Press.

Forbes, Daniel P. 1998. Measuring the Unmeasurable: Empirical Studies of Nonprofit Organization Effectiveness from 1977 to 1997. *Nonprofit and Voluntary Sector Quarterly* 27 (2):183-202.

Garrett, Geoffrey R., Daniel Kelemen, and Heiner Schulz. 1998. The European Court of Justice, National Governments and Legal Integration in the European Union. *International Organization* 52 (1):149-176.

Gartzke, Erik, Emilie M. Hafner-Burton, and Jana von Stein. 2008. International Organizations Count: What Statistics Tell Us About IOs. *Journal of Conflict Resolution (Special Issue)* April.

Gutner, Tamar L. 2002. *Banking on the Environment: Multilateral Development Banks and Their Environmental Performance in Central and Eastern Europe*. Cambridge, Mass.: MIT Press.

———. 2005. Explaining the Gaps between Mandate and Performance: Agency Theory and World Bank Environmental Reform. *Global Environmental Politics* 5 (2):10-37.

Haas, Peter M., Robert O. Keohane, and Marc A. Levy. 1993. *Institutions for the Earth: Sources of Effective International Environmental Protection*. Cambridge, Mass.: MIT Press.

Haftel, Yoram Z., and Alexander Thompson. 2006. The Independence of International Organizations. *Journal of Conflict Resolution* 50 (2):253-275.

Hawkins, Darren G., and Wade Jacoby. 2006. How Agents Matter. In *Delegation and Agency in International Organizations*, edited by D. G. Hawkins, D. A. Lake, D. Nielson and M. Tierney. New York: Cambridge University Press.

Hawkins, Darren G., David A. Lake, Daniel Nielson, and Michael Tierney. 2006. *Delegation and Agency in International Organizations*. New York: Cambridge University Press.

Held, David. 1995. *Democracy and the Global Order: from the Modern State to Cosmopolitan Governance*. Stanford, Calif.: Stanford University Press.

Helm, Carsten, and Detlef Sprinz. 2000. Measuring the Effectiveness of International Environmental Regimes. *Journal of Conflict Resolution* 44 (5):630-652.

Howard, Lise Morjâe. 2008. *UN Peacekeeping in Civil Wars*. New York: Cambridge University Press.

Johns, Leslie. 2007. A Servant of Two Masters: Communication and the Selection of International Bureaucrats. *International Organization* 61:245-275.

Kahler, Miles. 2001. *Leadership Selection in the Major Multilaterals*. Washington, DC: Institute for International Economics.

Kaplan, Robert S. 2001. Strategic Performance Measurement and Management in Nonprofit Organizations. *Nonprofit Management and Leadership* 11 (3):353-370.

Kay, David A., and Harold K. Jacobson. 1983. *Environmental Protection: the International Dimension*. Totowa, N.J.: Allanheld, Osmun & Co.

Kelley, Janet M. 2003. Citizen Satisfaction and Administrative Performance Measures: Is There Really a Link? *Urban Affairs Review* 38 (6):855-866.

Kirlin, John J. 2001. Big Questions for a Significant Public Administration. *Public Administration Review* 61 (2):140-43.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2004. *The Rational Design of International Institutions*. New York: Cambridge University Press.

Levy, Frank, Arnold J. Meltsner, and Aaron B. Wildavsky. 1974. *Urban Outcomes: Schools, Streets, and Libraries*. Berkeley: University of California Press.

Lipson, Michael. 2007. Peacekeeping: Organized Hypocrisy? *European Journal of International Relations* 13 (1):5-34.

Markus, M. Lynne, and Jeffrey Pfeffer. 1983. Power and the Design and Implementation of Accounting and Control Systems. *Accounting, Organizations, and Society* 8 (2/3):205-218.

Martens, Bertin. 2002. Introduction. In *The Institutional Economics of Foreign Aid*, edited by B. Martens, U. Mummert, P. Murrell and P. Seabright. New York: Cambridge University Press.

Mingst, Karen, and Craig Warkentin. 1996. What Difference Does Culture Make in Multilateral Negotiations? *Global Governance* 2 (2):169-188.

Mitchell, Ronald B. 1994. *Intentional Oil Pollution at Sea: Environmental Policy and Treaty Compliance*. Cambridge, Mass.: MIT Press.

Moe, Terry M. 1989. The Politics of Bureaucratic Structure. In *Can the Government Govern?*, edited by J. E. Chubb and P. E. Peterson. Washington, D.C.: Brookings Institution.

———. 2005. Power and Political Institutions. *Perspectives on Politics* 3 (2):215-233.

Nakamura, Robert T., and Frank Smallwood. 1980. *The Politics of Policy Implementation*. New York: St. Martin's Press.

NAPA. 2002. *Performance Indicators: Getting Started*. Washington, D.C.: National Academy of Public Administration.

Paris, Roland. 2003. Peacekeeping and the Constraints of Global Culture. *European Journal of International Relations* 9 (3):441-473.

Perrow, Charles. 1961. The Analysis of Goals in Complex Organizations. *American Sociological Review* 26 (6):854-866.

———. 1986. *Complex organizations: a critical essay*. 3rd ed. New York: Random House.

Pollack, Mark A. 2003. *The Engines of European Integration: Delegation, Agency, and Agenda Setting in the EU*. New York: Oxford University Press.

Powell, Walter W. 1991. Expanding the Scope of Institutional Analysis. In *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. DiMaggio. Chicago: University of Chicago Press.

Reinalda, Bob, and Bertjan Verbeek. 2004. The issue of decision making within international organizations. In *Decision Making Within International Organizations*, edited by B. Reinalda and B. Verbeek. New York: Routledge.

Shanks, Cheryl, Harold K. Jacobson, and Jeffrey H. Kaplan. 1996. Inertia and Change in the Constellation of International Governmental Organizations, 1981-1992. *International Organization* 50 (4):593-627.

Simmons, Beth A. 2000. International Law and State Behavior: Commitment and Compliance in International Monetary Affairs. *American Political Science Review* 94 (4):819-835.

Simon, Herbert A., Donald W. Smithburg, and Victor A. Thompson. 1973. *Public Administration*. New York: Alfred A. Knopf.

Thompson, Alexander. 2006. Screening Power: International Organizations as Informative Agents. In *Delegation and Agency in International Organizations*, edited by D. G. Hawkins, D. A. Lake, D. Nielson and M. Tierney. New York: Cambridge University Press.

———. 2008. Principal Problems: the Rise and Fall of UN Weapons Inspections in Iraq and Beyond. Unpublished manuscript: Department of Political Science, Ohio State University.

Underdal, Arild. 2002. One Question, Two Answers. In *Environmental Regime Effectiveness: Confronting Theory with Evidence*, edited by E. L. Miles, A. Underdal, S. Andresen, J. Wettestad, J. B. Skjaerseth and E. M. Carlin. Cambridge, MA: MIT Press.

Underdal, Arild, and Oran R. Young. 2004. *Regime Consequences: Methodological Challenges and Research Strategies*. New York: Kluwer Academic Publishers.

Vaubel, Roland. 2006. Principal-Agent Problems in International Organizations. *Review of International Organizations* 1 (2):125-138.

Vreeland, James R. 2006. IMF program compliance: Aggregate index versus policy specific research strategies. *Review of International Organizations* 1 (4):359-378.

Weaver, Catherine. 2007. The World's Bank and the Bank's World. *Global Governance* 13 (4):493-512.

Wendt, Alexander. 2001. Driving with the Rearview Mirror: on the Rational Science of Institutional Design. *International Organization* 55 (4).

Wilkinson, Rorden. 2008.

Young, Oran R. 1999. *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. Cambridge, Mass.: MIT Press.

Zweifel, Thomas D. 2006. *International Organizations and Democracy: Accountability, Politics, and Power*. Boulder, Colo.: Lynne Rienner.