

The politics of IO performance: A framework

Tamar Gutner · Alexander Thompson

Received: 19 July 2010 / Revised: 19 July 2010 / Accepted: 19 July 2010 /
Published online: 31 July 2010
© Springer Science+Business Media, LLC 2010

Abstract Some international organizations (IOs) are subject to constant criticism for producing poor results while others are praised for accomplishing difficult tasks despite political and resource constraints. Indeed, IO performance varies substantially over time and across tasks, and yet the international relations literature has devoted little attention to why this occurs. This article provides a framework for studying IO performance. After addressing some of the distinct challenges of conceptualizing and analyzing performance in the context of IOs, we discuss the tradeoffs of using different performance metrics—from process indicators to outcome indicators—and present a typology of factors that influence performance. Finally, we discuss research strategies for those interested in studying performance rigorously. The policy relevance of studying IO performance is clear: only if we understand why some IOs perform better than others can we begin to improve their performance in a systematic way. As many organizations come under pressure to reform, while at the same time taking on new and more complicated tasks, scholars should be actively engaged in debates surrounding IO performance and its role in effective governance at the international level.

Keywords International organizations · International relations · Institutions · Performance

JEL Codes F53 · F55

1 Introduction

International organizations (IOs) are essential but controversial actors in world politics today. They are expected to rebuild war-torn societies, reduce extreme

T. Gutner (✉)

School of International Service, American University, 4400 Massachusetts Avenue, N.W.,
Washington, DC 20016, USA
e-mail: tgutner@american.edu

A. Thompson

Department of Political Science, Ohio State University, 2140 Derby Hall, 154 N. Oval Mall,
Columbus, OH 43210, USA

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 domestic 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 literature is underdeveloped in its ability to explain why this occurs—and why IOs sometimes perform quite well. Why do some IOs perform better than others? Why does IO performance vary, both over time and across issues? How can performance be improved? 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. This article provides a foundation for engaging in this exercise. Along with the other contributors to this volume, we hope to improve our ability to explore international organization performance (IOP) in ways that both contribute to theory and inform policy discussions.

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 decision making, 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,¹ and thus our ability to understand performance—what it is and where it comes from—is crucial.

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.” 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 institutions. Obviously, distinct and specific criteria must be used to evaluate individual 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. Nevertheless, we believe it is possible to provide some

¹ For more discussion of performance as a source of IO legitimacy, see Cronin and Hurd (2008).

guidance for thinking about performance in general, analytical terms. At a time when many IOs face new challenges and require rethinking and reform, it is critical that scholars do more than sit passively on the sidelines.

This article proposes a 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. There are tradeoffs associated with these conceptualization and measurement choices; we discuss their strengths and weaknesses and highlight the need to tailor an approach to fit particular organizations and issues. 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 politics), 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 factors. A concluding section introduces the other contributions to this special issue, which apply and reflect on our framework to explore IOP in a range of important empirical cases. We end with some tentative lessons and conjectures derived from the volume as a whole.

2 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 et al. 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 can clearly impact an IO’s performance in numerous ways. 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). We therefore 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 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 et al. 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 critical question of whether a given design leads to better performance and improved outcomes (Wendt 2001). We incorporate insights from the design and delegation literatures but view these characteristics of IOs as independent variables that help explain IOP rather than as dependent variables.

Finally, the 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)?

The regime effectiveness literature offers suggestions on defining and measuring institutional effects and has grappled with many of the methodological problems discussed here.² However, 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, as Abbott and Snidal (2010) discuss in this volume. Second, most work on regime effectiveness has focused on environmental problems 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. IOs, like most other public 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).

² For an interesting debate on some of these issues, see Hovi et al. (2003) and Young (2001).

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.

3 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 include some sort of centralized administrative apparatus with a permanent staff.³ IOs are thus more independent and formalized than other international bodies (Bradley and Kelley 2008; Abbott and Snidal 1998). Performance reflects the behavior of two sets of actors, the member states and the staff. The causal influences on 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 is 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 outcome 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 may involve breaking long-term objectives down into more specific medium-term objectives, and then developing (usually quantifiable) performance measures that are used to determine the extent to which the objectives are accomplished (Radin 2006). This traditional technique 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

³ For a similar definition, see Shanks et al. (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.

(Kaplan 2001; Simon et al. 1973). In the public administration literature, scholars also note that performance measures may be used in areas that go well beyond basic evaluation, and provide public managers with data that can also help promote an agency, celebrate its accomplishments, motivate staff, and control subordinates (Behn 2003).

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 complex. 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.

4 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 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 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 (Glennon 2003; Thompson 2009). Alternatively, the Security Council may be judged as a failure in its governance function of maintaining international peace and security but be more successful in its role as a “loose concert” of powerful states (Bosco 2009). The eye of the beholder 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 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). Lipson (2010) points out in this issue 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 starkly opposed perceptions on the performance of virtually any major IO makes it even more important for scholars to offer better ways of analyzing 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 staff 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 (2010) points out the “paradox” of self-evaluation, in that independent evaluation units within IOs are supposed to improve

most intuitive way to evaluate institutional effects. If outcomes can be measured in a way that is standardized across cases, this approach holds the promise of allowing for comparison across IOs and hence for generalization (Hovi et al. 2003).

However, 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. Moreover, many IOs perform only limited functions, such as coordination and information-gathering, or operate in issue-areas where multiple institutions combine to supply governance. Such organizations should not be held responsible for whether the larger problem is solved. Finally, some IOs have multiple goals, which complicates the use of outcome metrics. 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 impact on 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 et al. 1993). The public policy literature refers to these as the “outputs” or “intermediate outcomes” of government programs (Levy et al. 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 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).

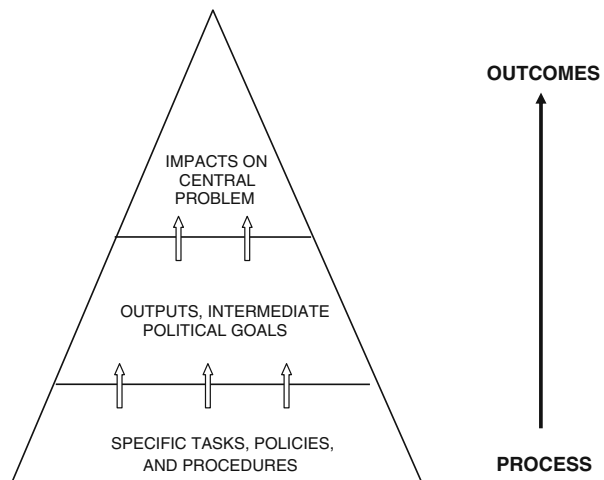
Figure 2 helps us understand these analytical tradeoffs by portraying the various stages of IO performance as a pyramid. This illustrates places at which performance

⁴ For more discussion of this terminology, though with somewhat different categories, see Underdal (2002) and Easton (1965).

may be observed and assessed, rather than offering one overarching set of independent or dependent variables. 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. Ideally, 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 compliance with trade rules. 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.

Fig. 2 Pyramid of performance:
from process to outcomes



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.

6 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 et al. 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 staff but the result of problems emanating from member states—what Thompson (2006) refers to as “principal problems.” 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 member states have delegated conflicting and complex tasks that are difficult for the institution to implement. In sum, this first approach views IO performance, 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 performance. Barnett and Finnemore’s (2004) work on the importance of bureaucratic culture 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 Paul Wolfowitz' effect on staff morale at the World Bank illustrates (Weaver 2008). 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 material interests 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. 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 increase in multilateral peacekeeping in the 1990s, and Ian Hurd (2007) explains the value of the Security Council in terms of its widely perceived legitimacy. In some cases, IOs may perform poorly because their missions do not reflect a clear consensus among states of what normative principles should be pursued or what underlying problem needs to 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). 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 an efficient set of procedures for carrying out its work, but still fail if it is working in a situation where there is instability, weak capacity, corruption, or a lack of consent from relevant parties (Gutner 2005; Howard 2008).

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. These factors can bear on performance at the level of process or outcomes.

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). More generally, the external political world of states and the internal bureaucratic setting of the staff interact, and the two sets of actors exist in a mutually dependent relationship (Reinalda and Verbeek 2004).

Table 1 The sources of performance

	Internal	External
Social	<ul style="list-style-type: none"> –Organizational culture –Leadership deficit 	<ul style="list-style-type: none"> –Competing norms –Lack of consensus on problem
Material	<ul style="list-style-type: none"> –Inadequate staffing, resources –Bureaucratic/career self-interest 	<ul style="list-style-type: none"> –Power politics among member states –Incoherent mandates –On-the-ground constraints

Along these lines, some argue for more attention to ways in which 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 Catherine Weaver's (2007) 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. 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.

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. 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 results (Wilkinson 2006). There may also be cases where member states' interests are aligned, yet something happens inside the bureaucracy to throw performance off track, such as a mismatch between member-state goals and the incentives of bureaucrats (Pollack and Hafner-Burton 2010).

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 behavior seriously. We build on the same premises.

7 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 the mission at a specific point in time. Scholars can narrow ambiguous or contested 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 the first stage by explaining how they capture performance results of interest.
- 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 (Fig. 2): narrow administrative functions, intermediate goals and outputs, 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 and describe the mechanisms by which they shape IOP. We should be open to the possibility that a combination of factors—social and material, internal and external—is driving outcomes.

Two more general methodological 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, as Hyde (2009) has done with democracy assistance programs. When this approach is impractical and data are not available, 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 can be helpful.

⁵ Bernauer (1995) makes a similar point in the context of regime effectiveness. Some of the contributions to Haas et al. (1993) are guided by similar counterfactual strategies.

8 Contributions to the Volume: Findings and Conjectures

The other contributors to this special issue draw on the conceptual and analytical framework presented in this article to advance the study of IO performance. Most investigate IOP in the context of a specific issue area or IO and address one or more of the following questions. What is the best way to measure IO performance in the empirical context under study? Does the organization in question exhibit good or bad performance? What explains this performance outcome? Some authors also speculate on how performance could be improved for the IO or IOs they have examined. In the process of conducting their research, the contributors have been encouraged to reflect critically on our framework and to offer alternative concepts, propositions, and research strategies. The result is a set of papers that are rich and eclectic but also sufficiently coherent to speak to each other and generate both tentative conclusions and conjectures for future research.

Michael Lipson focuses on some of the challenges facing scholars and policy-makers who seek to evaluate the United Nations' performance in peacekeeping. The author draws from the sociology literature on organizations to show how various forms of ambiguity contribute to the difficulty of defining and assessing the performance of UN peace operations. Indeed, he views the relationship between process performance and outcomes in this area as "irreducibly ambiguous," which has very real implications for efforts to improve peacekeeping. Analyzing performance at the level of process rather than outcomes, Lipson argues the prevailing "results-based budgeting" approach of the UN to its internal management has failed to improve performance because it is poorly suited to the inherent ambiguities of the organizational environment.

Mark Pollack and Emilie Hafner-Burton examine through a rationalist lens the ability of the EU to implement the mandates of gender mainstreaming and environmental policy integration. These are examples of what the authors call "cross-cutting mandates" that are supposed to be addressed or "mainstreamed" across the EU. The authors make a case for relying on policy *outputs* as the best way to measure performance, by examining the factors that shape how well Commission Directorates-General and services develop policies aimed at addressing gender inequality and environmental quality. They argue that the incentives facing these officials greatly influence whether such mainstreaming exercises succeed or fail. In particular, the Commission's emphasis on soft incentives, such as persuasion and socialization, results in only partial progress in promoting these goals. Hard, material incentives, either positive or negative, are more effective at motivating the relevant officials, which results in the development of stronger gender and environmental policies.

Rather than focusing on a specific IO, Kenneth Abbott and Duncan Snidal explore how IOs in general can enhance their performance in the area of regulatory standard setting by reaching out to a combination of governmental and private actors. They describe an emerging world of Transnational New Governance, where the most valuable role of IOs is in the "orchestration" of global governance networks in areas such as the environment, human rights and workers' rights. Because IOs are only one actor in this setting and serve a limited role, a variety of performance measures, mostly focused on elements of process rather than outcomes, must be used

to assess their contribution. Orchestration offers the potential to improve IO performance in the long run by bringing various stakeholders together to agree on goals and to scrutinize the IO's activities, and by encouraging IOs to develop a wider range of regulatory techniques that are less dependent on states.

Manfred Elsig examines the WTO and notes that its goals are variegated and contested, making it difficult to establish a "baseline" against which to measure performance. He settles on four important functions of the organization—negotiations, dispute settlement, regime management, and technical assistance—and outlines the complex principal-agent relationships and "institutional milieu" that shapes WTO performance. His main focus is on the ability of the Secretariat to translate process performance into outcome performance in the context of these complex political relationships and despite the constraints of a "member-driven" organization, factors that often limit the autonomy of the Secretariat and the availability of high quality information across WTO tasks.

Catherine Weaver explores some of the problems inherent in IO efforts to set up mechanisms for internal evaluation. Using considerable primary research, she outlines the history of the IMF's Independent Evaluation Office (IEO) and explores some of the political challenges it faces in creating strong mechanisms for self-assessment. These challenges include how to maintain independence, how to develop metrics to assess the Fund's performance, how to maintain credibility in the eyes of internal and external audiences (who may disagree over the organization's most important goals), and how to foster a culture of learning within the IMF. Ultimately, the IEO's ability to do its own work well has consequences for its ability to influence the IMF's broader performance, including the social and economic outcomes that result from its lending programs.

Taken together, these contributions to the special issue reinforce and sharpen many of the points made in this article. They also bring to light new analytical challenges for the study of IO performance and suggest conjectures—deserving of future research—regarding the sources of good and bad performance. We summarize some of the most interesting findings here.

Baselines and Beholders Good performance is often difficult to judge because many IOs have multiple objectives and because those who control and are affected by IOs disagree over what constitutes success and at which level of analysis. This point is made most forcefully by Lipson, in his study of UN peacekeeping, and by Elsig, in his study of the WTO. In both cases there are multiple stakeholders and considerable disagreement over which organizational goals are paramount. Abbott and Snidal note that firms, NGOs, and government all have distinct preferences when it comes to the governance of regulatory standards. Because most IOs face an "eye of the beholder" problem when it comes to evaluating performance, the analyst is obliged to be specific about what organizational objectives are to be analyzed. All of the authors in this volume have engaged in this baseline-establishing exercise in one way or another, and it is an important part of any research strategy for studying IO performance.

Performance Under Anarchy The relative anarchy of the international political system has important implications for the study of the IO performance. First, it

means that IOs often play a limited role when it comes to global governance, as states and private actors share responsibility. In Abbott and Snidal's study, IOs are "orchestrating" quite decentralized activities rather than providing top-down regulation. The WTO provides another example, where the Secretariat is primarily charged with facilitating cooperation among states, a role outlined in detail in Elsig's contribution. Second, one empirical pattern that emerges from the articles in this special issue is that good internal performance by IO bureaucracies is often threatened by the interests and interventions of outside actors, especially states but also non-governmental actors. Weaver, Elsig, and Lipson all offer clear examples of the negative consequences of outside interference, which can hamper process performance or prevent good process performance from translating into good outcome performance. Those focused on reforming the internal characteristics of IOs should keep in mind that, in many cases, these bureaucracies are only as effective as external actors allow them to be.

Secretariat Autonomy The articles in this volume show that IO bureaucracies clearly matter and have important roles to play in contributing to their organizations' performance. Abbott, Snidal and Elsig argue that bureaucracies will perform their functions—usually related to process and outputs—better if they are given some degree of autonomy. Abbott and Snidal argue explicitly that some independence is necessary for effective orchestration by IOs, and Elsig notes that the Secretariat's Trade Policy Reviews have gained in quality in the absence of micro-managing by states. A related lesson of Weaver's study is that effective self-evaluation requires some political insulation for the office in question. Another observation is that IO bureaucracies seem to have the most autonomy when they perform micro-functions related to process but are more constrained when they engage in macro-functions that are more politically visible. An important policy puzzle, therefore, is how to provide secretariats with more autonomy across a range of activities (moving up the performance pyramid) without threatening government concerns over control and sovereignty.

Bureaucratic Incentives Poor performance is likely to occur when the incentives of IO staff do not match the incentives of the IO's leadership (in terms of both internal management and member-state governments). This creates multiple principal-agent problems when it comes to guiding staff behavior, as Elsig notes. One implication is that IO leaders must think seriously about the incentives of the staff to align their behavior with organizational goals. This is especially striking in the studies by Lipson and Weaver, showing that credible self-evaluation is highly sensitive to the professional incentives and organizational culture of staff members, who often view outside interference with suspicion. For Pollack and Hafner-Burton the key issue is what types of incentives will mobilize IO bureaucrats to take cross-cutting mandates seriously, and their argument that "hard" incentives are more likely to change staff behavior than softer forms of socialization and persuasion is worth exploring beyond the European Commission case they present.

Difficult Mandates One major impediment to good performance is the existence of hopelessly complex, ambitious, or ambiguous mandates. The authors in this volume

provide multiple examples of IO staff saddled with mandates that are difficult to achieve. Weaver explains the tension created when independent evaluation bodies are asked to evaluate an IO frankly but also to prove to external observers the IO is worthy of support. Lipson points to the inappropriate evaluation and management techniques imposed on UN officials, while Pollack and Hafner-Burton warn that the increasing number of “mainstreaming” mandates threatens to overwhelm EU officials, especially when the pursuit of these mandates entails tradeoffs and contradictions. These findings are consistent with previous research on the problems with many IO mandates and suggest that IO reform efforts should focus on the clarification and rationalization of mandates promulgated by member-states.

The Interdependence of Global Governors The ability of IOs to succeed often depends on the activities of other governance actors and institutions that co-exist in a given issue-area (Avant et al. 2010). Sometimes a division of labor emerges where various actors play complementary roles—Abbott and Snidal point to this possibility—but other times there is conflict between the activities of an IO and its competitors (See Oberthür and Gehring 2006). While global cooperation and coordination sound good in principle, in practice such efforts may complicate already-difficult issues of control, outlined here by Elsig, and weaken mechanisms of accountability (Gutner 2010). Who is responsible when everyone is acting together and things go wrong?

The Forest or the Trees? Often our conclusions regarding performance are shaped by the power of our analytical microscope. One observer looking at the detailed inner-workings of an IO might come to different conclusions about its performance than another looking at its broader role in world politics. An assessment of UNEP focused on its internal design, resources and leadership concludes that its performance record is mixed at best (Ivanova 2010), whereas Abbott and Snidal (2010) offer a relatively favorable assessment of UNEP’s broader role as a global orchestrator. Similarly, Weaver offers a relatively positive assessment of the IEO, a subunit of the IMF, whereas assessments of the broader economic effects of IMF lending tend to be more negative (Stiglitz 2002; Vreeland 2003). In the end, both exercises are equally valuable and help us understand different aspects of performance.

Process Versus Outcomes As we discuss at length above, performance can be conceptualized and measured at the level of process or the level of outcomes, or somewhere in between. For the most part, the authors in this volume focus on process performance, and note that studying broader organizational impacts presents difficult methodological problems associated with defining and measuring the relevant goals and linking IO activities causally to them. Notably, the studies in this volume are mostly qualitative in nature, an approach that lends itself to analyzing the details of process and output performance. Quantitative studies of IO effectiveness tend to select objective outcomes and impacts for purposes of measurement and comparability, and they offer the advantages of generalization and more sophisticated controls (See, e.g., Flores and Nooruddin 2009; Steinwand and Stone 2008). Because they both add value to the study of IO performance, qualitative and

quantitative studies should work in tandem according to their comparative advantages and the specific empirical puzzles of interest.

Our goal with this special issue is not to provide or test a single viewpoint but rather to provide a set of compelling papers on IO performance that speak to a common set of theoretical, empirical and methodological questions. More broadly, our goal is to catalyze the emerging research agenda on this topic in the field of IO, an agenda with an obvious normative dimension. In the face of major global challenges, there is an increased need for IOs to better address old problems and to take on a growing list of new ones. The ability of IOs to perform well in these efforts has important implications for the shape and success of global governance—and, ultimately, for human welfare—in the decades ahead.

Acknowledgements This project received generous support from an International Studies Association Venture Research Workshop Grant. We are grateful to Miles Kahler and Bob Keohane for providing comments on many of the papers, and to Susan Hyde and Rorden Wilkinson for their role in shaping the project from its earliest days. For comments on this article, we thank Bruno Frey, Erin Graham, Bob Keohane, Sebastian Oberthür, Jordan Tama, Mike Tierney, and the international group of scholars who actively contributed to our ISA workshop. We are indebted to the other authors in this special issue for their intellectual engagement and valuable suggestions along the way. Alex Thompson thanks the Mershon Center at Ohio State for financial support.

References

- Abbott, K. W., & Snidal, D. (1998). Why states act through formal international organizations. *Journal of Conflict Resolution*, 42(1), 3–32.
- Abbott, K. W., & Snidal, D. (2010). International regulation without international government: improving IO performance through orchestration. *Review of International Organizations*, 5(3).
- Alter, K. J. (1998). Who are the “masters of the treaty”? European governments and the European court of justice. *International Organization*, 52(1), 121–147.
- Alter, K. J., & Meunier, S. (2008). The politics of international regime complexity. *Perspectives on Politics*, 7(1), 13–24.
- Avant, D. D., Finnemore, M., & Sell, S. K. (Eds.). (2010). *Who governs the globe*. New York: Cambridge University Press.
- Barnett, M. N. (2002). *Eyewitness to a genocide: The United Nations and Rwanda*. Ithaca: Cornell University Press.
- Barnett, M. N., & Finnemore, M. (1999). The politics, power, and pathologies of international organizations. *International Organization*, 53(4), 699–732.
- Barnett, M. N., & Finnemore, M. (2004). *Rules for the world: International organizations in global politics*. Ithaca: Cornell University Press.
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606.
- Bernauer, T. (1995). The effect of international environmental institutions: how we might learn more. *International Organization*, 49(2), 351–377.
- Bosco, D. L. (2009). *Five to rule them all: The UN security council and the making of the modern world*. New York: Oxford University Press.
- Bradley, C. A., & Kelley, J. G. (2008). The concept of international delegation. *Law and Contemporary Problems*, 71(1), 1–36.
- Buchanan, A., & Keohane, R. O. (2006). The legitimacy of global governance institutions. *Ethics and International Affairs*, 20(4), 405–437.
- Cooley, A., & Ron, J. (2002). The NGO scramble: organizational insecurity and the political economy of transnational aid. *International Security*, 27(1), 5–39.

- Cronin, B., & Hurd, I. (2008). *The UN security council and the politics of international authority*. New York: Routledge.
- Dahl, R. A. (1999). Can international organizations be democratic? A skeptic's view. In I. Shapiro & C. Hacker-Cordón (Eds.), *Democracy's edges*. New York: Cambridge University Press.
- Easton, D. (1965). *A systems analysis of political life*. New York: Wiley.
- Elsig, M. (2010). The world trade organization at work: performance in a member-driven milieu. *Review of International Organizations*, 5(3).
- Finnmore, M. (1996). *National interests in international society*. Ithaca: Cornell University Press.
- Flores, T. E., & Nooruddin, I. (2009). Financing the peace: evaluation world bank post-conflict assistance programs. *Review of International Organizations*, 4, 1–27.
- Forbes, D. P. (1998). Measuring the unmeasurable: empirical studies of nonprofit organization effectiveness from 1977 to 1997. *Nonprofit and Voluntary Sector Quarterly*, 27(2), 183–202.
- Fortna, V. P. (2008). *Does peacekeeping work? Shaping belligerents' choices after civil war*. Princeton: Princeton University Press.
- Garrett, G. R., Kelemen, D., & Schulz, H. (1998). The European court of justice, national governments and legal integration in the European union. *International Organization*, 52(1), 149–176.
- Gartzke, E., Hafner-Burton, E. M., & von Stein, J. (2008). International organizations count: what statistics tell us about IOs. *Journal of Conflict Resolution (Special Issue)*, 52(2).
- Glennon, M. (2003). Why the security council failed. *Foreign Affairs*, 82(3), 16–35.
- Gutner, T. L. (2005). Explaining the gaps between mandate and performance: agency theory and world bank environmental reform. *Global Environmental Politics*, 5(2), 10–37.
- Gutner, T. (2010). When 'doing good' does not: The IMF and the millennium development goals. In D. D. Avant, M. Finnmore, & S. K. Sell (Eds.), *Who governs the globe?* New York: Cambridge University Press.
- Haas, P. M., Keohane, R. O., & Levy, M. A. (1993). *Institutions for the earth: Sources of effective international environmental protection*. Cambridge: MIT Press.
- Haftel, Y. Z., & Thompson, A. (2006). The independence of international organizations. *Journal of Conflict Resolution*, 50(2), 253–275.
- Hawkins, D. G., & Jacoby, W. (2006). How agents matter. In D. G. Hawkins, D. A. Lake, D. Nielson, & M. Tierney (Eds.), *Delegation and agency in international organizations*. New York: Cambridge University Press.
- Hawkins, D. G., Lake, D. A., Nielson, D., & Tierney, M. (2006). *Delegation and agency in international organizations*. New York: Cambridge University Press.
- Held, D. (1995). *Democracy and the global order: From the modern state to cosmopolitan governance*. Stanford: Stanford University Press.
- Helm, C., & Sprinz, D. (2000). Measuring the effectiveness of international environmental regimes. *Journal of Conflict Resolution*, 44(5), 630–652.
- Hovi, J., Sprinz, D., & Underdal, A. (2003). The Oslo-Potsdam solution to measuring regime effectiveness: critique, response, and the road ahead. *Global Environmental Politics*, 3(3), 74–96.
- Howard, L. M. (2008). *UN peacekeeping in civil wars*. New York: Cambridge University Press.
- Hurd, I. (2007). *After anarchy: Legitimacy and power in the united nations security council*. Princeton: Princeton University Press.
- Hyde, S. D. (2009). Experimenting in democracy promotion: international observers and the 2004 presidential elections in Indonesia. *Perspectives on Politics*, 8(2), 511–527.
- Ivanova, M. (2010). UNEP in global environmental governance: design, leadership, location. *Global Environmental Politics*, 10(1), 30–59.
- Johns, L. (2007). A servant of two masters: communication and the selection of international bureaucrats. *International Organization*, 61, 245–275.
- Kahler, M. (2001). *Leadership selection in the major multilaterals*. Washington: Institute for International Economics.
- Kaplan, R. S. (2001). Strategic performance measurement and management in nonprofit organizations. *Nonprofit Management & Leadership*, 11(3), 353–370.
- Kay, D. A., & Jacobson, H. K. (1983). *Environmental protection: The international dimension*. Totowa: Allanheld, Osmun & Co.
- Kelley, J. M. (2003). Citizen satisfaction and administrative performance measures: is there really a link? *Urban Affairs Review*, 38(6), 855–866.
- Keohane, R. O. (1984). *After hegemony: Cooperation and discord in the world political economy*. Princeton: Princeton University Press.

- Kirlin, J. J. (2001). Big questions for a significant public administration. *Public Administration Review*, 61(2), 140–143.
- Koremenos, B., Lipson, C., & Snidal, D. (2004). *The rational design of international institutions*. New York: Cambridge University Press.
- Levy, F., Meltzer, A. J., & Wildavsky, A. B. (1974). *Urban outcomes: Schools, streets, and libraries*. Berkeley: University of California Press.
- Lipson, M. (2007). Peacekeeping: organized hypocrisy? *European Journal of International Relations*, 13(1), 5–34.
- Lipson, M. (2010). Performance under ambiguity: international organization performance in UN peacekeeping. *Review of International Organizations*, 5(3).
- Lyne, M. M., Nielson, D. L., & Tierney, M. J. (2006). Who delegates? Alternative models of principals in development aid. In D. G. Hawkins, D. A. Lake, D. L. Nielson, & M. J. Tierney (Eds.), *Delegation and agency in international organizations*. New York: Cambridge University Press.
- Markus, M. L., & Pfeffer, J. (1983). Power and the design and implementation of accounting and control systems. *Accounting, Organizations and Society*, 8(2/3), 205–218.
- Martens, B. (2002). Introduction. In B. Martens, U. Mummert, P. Murrell, & P. Seabright (Eds.), *The institutional economics of foreign aid*. New York: Cambridge University Press.
- Mingst, K., & Warkentin, C. (1996). What difference does culture make in multilateral negotiations? *Global Governance*, 2(2), 169–188.
- Mitchell, R. B. (1994). *Intentional oil pollution at sea: Environmental policy and treaty compliance*. Cambridge: MIT Press.
- Moe, T. M. (1989). The politics of bureaucratic structure. In J. E. Chubb & P. E. Peterson (Eds.), *Can the government govern?* Washington, D.C.: Brookings Institution.
- Moe, T. M. (2005). Power and political institutions. *Perspectives on Politics*, 3(2), 215–233.
- Nakamura, R. T., & Smallwood, F. (1980). *The politics of policy implementation*. New York: St. Martin's.
- NAPA. (2002). *Performance indicators: Getting started*. Washington, D.C.: National Academy of Public Administration.
- Oberthür, S., & Gehring, T. (2006). *Institutional interaction in global environmental governance: Synergy and conflict among international and EU policies*. Cambridge: MIT Press.
- Paris, R. (2003). Peacekeeping and the constraints of global culture. *European Journal of International Relations*, 9(3), 441–473.
- Perrow, C. (1961). The analysis of goals in complex organizations. *American Sociological Review*, 26(6), 854–866.
- Perrow, C. (1986). *Complex organizations: A critical essay* (3rd ed.). New York: Random House.
- Pollack, M. A. (2003). *The engines of European integration: Delegation, agency, and agenda setting in the EU*. New York: Oxford University Press.
- Pollack, M. A., & Hafner-Burton, E. M. (2010). Mainstreaming international governance: the environment, gender, and IO performance in the European union. *Review of International Organizations*, 5(3).
- Powell, W. W. (1991). Expanding the scope of institutional analysis. In W. W. Powell & P. DiMaggio (Eds.), *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press.
- Radin, B. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values, public management and change series*. Washington, D.C.: Georgetown University Press.
- Reinalda, B., & Verbeek, B. (2004). The issue of decision making within international organizations. In B. Reinalda & B. Verbeek (Eds.), *Decision making within international organizations*. New York: Routledge.
- Shanks, C., Jacobson, H. K., & Kaplan, J. H. (1996). Inertia and change in the constellation of international governmental organizations, 1981–1992. *International Organization*, 50(4), 593–627.
- Simmons, B. A. (2000). International law and state behavior: commitment and compliance in international monetary affairs. *American Political Science Review*, 94(4), 819–835.
- Simon, H. A., Smithburg, D. W., & Thompson, V. A. (1973). *Public administration*. New York: Alfred A. Knopf.
- Steinwand, M. C., & Stone, R. W. (2008). The international monetary fund: a review of the recent evidence. *Review of International Organizations*, 3, 123–149.
- Stiglitz, J. E. (2002). *Globalization and its discontents* (1st ed.). New York: W. W. Norton.
- Thompson, A. (2006). Principal problems: The rise and fall of UN weapons inspections in Iraq. Unpublished manuscript: Department of Political Science, Ohio State University.
- Thompson, A. (2009). *Channels of power: The UN security council and U.S. statecraft in Iraq*. Ithaca: Cornell University Press.

- Underdal, A. (2002). One question, two answers. In E. L. Miles, A. Underdal, S. Andresen, J. Wettstad, J. B. Skjaereth, & E. M. Carlin (Eds.), *Environmental regime effectiveness: Confronting theory with evidence*. Cambridge: MIT Press.
- Underdal, A., & Young, O. R. (2004). *Regime consequences: Methodological challenges and research strategies*. New York: Kluwer Academic.
- Vaubel, R. (2006). Principal-agent problems in international organizations. *Review of International Organizations*, 1(2), 125–138.
- Vreeland, J. R. (2003). *The IMF and economic development*. New York: Cambridge University Press.
- Vreeland, J. R. (2006). IMF program compliance: aggregate index versus policy specific research strategies. *Review of International Organizations*, 1(4), 359–378.
- Weaver, C. (2007). The world's bank and the bank's world. *Global Governance*, 13(4), 493–512.
- Weaver, C. (2008). *Hypocrisy trap: The world bank and the poverty of reform*. Princeton: Princeton University Press.
- Weaver, C. (2010). The politics of performance evaluation: independent evaluation at the international monetary fund. *Review of International Organizations*, 5(3).
- Wendt, A. (2001). Driving with the rearview mirror: on the rational science of institutional design. *International Organization*, 55(4), 1019–1049.
- Wilkinson, R. (2006). *The WTO: Crisis and the governance of global trade*. New York: Routledge.
- Young, O. R. (1999). *The effectiveness of international environmental regimes: Causal connections and behavioral mechanisms*. Cambridge: MIT Press.
- Young, O. (2001). Inferences and indices: evaluating the effectiveness of international environmental regimes. *Global Environmental Politics*, 1(1), 99–121.
- Zweifel, T. D. (2006). *International organizations and democracy: Accountability, politics, and power*. Boulder: Lynne Rienner.