Introduction

1

A Framework for Evaluating the Performance of International Courts and Tribunals

THERESA SQUATRITO, ORAN R. YOUNG, ANDREAS FOLLESDAL, AND GEIR ULFSTEIN

Introduction

Many analysts have noted a trend toward the legalization of international affairs in recent decades (Goldstein et al. 2000). While the extent of this development is hard to measure, one significant feature of legalization is unmistakable. We have witnessed a proliferation of international courts and tribunals (ICs).[1] Such bodies now operate globally and in several regions of the world; they play significant roles in the application and interpretation of many elements of international law. ICs address issues of regional integration, trade, and economic relations. They have become important elements in human rights systems in Europe, the Americas, and Africa. ICs are integral to the development of international criminal law and help to determine territorial and maritime boundaries.

There is significant variation among ICs regarding not only their mandates, but also the practices they have adopted and the effects of their rulings. International courts vary greatly with regard to their level of activity. Some courts are extremely busy institutions. The European Court of Human Rights (ECtHR) and the Court of Justice of the European Union (CJEU)[2] both issue hundreds of decisions annually. The ECtHR issued 916 judgments responding to 3,659 applications in 2013, while the CJEU issued 1322 decisions by judgment or order in 2013.[3] On the other hand, there are courts

[1] We use the abbreviation ICs rather than ICTs to avoid confusion with international criminal tribunals.
[2] Formerly known as the European Court of Justice (ECJ).
[3] The figures for the CJEU include the decisions of the General Court and the Court of Justice. For statistics on the ECtHR's judicial activity, see www.echr.coe.int/Pages/home.aspx?p=reports&c= (accessed July 27, 2017). For statistics on the judicial activity of the CJEU, see http://curia.europa.eu/jcms/jcms/Jo2_7000/ (accessed July 27, 2017).

like the International Court of Justice (ICJ) that hear only a handful of cases in any given year. ICs differ also in their practices pertaining to access, transparency and confidentiality, fact-finding, standards of review, methods of interpretation, and more.

Some ICs have become "islands" of litigation, generally addressing a narrower set of legal questions than their mandates cover. Approximately 90 percent of the cases on the docket of the Andean Tribunal of Justice, for example, pertain to intellectual property law, though the mandate of the Court is much broader (Helfer et al. 2009). On the other hand, there are courts that have exercised authority across the entire scope of their mandate; some have even expanded their jurisdiction. Similarly, some ICs have demonstrated judicial activism, while others have shown judicial restraint.

Several international courts have encountered serious problems, facing backlogs in their caseloads, backlash on the part of member states, or failure to become operational. The ECtHR, for example, is saddled with an overwhelming backlog of cases; it had a backlog of 99,900 applications in 2013. The Southern African Development Community (SADC) Tribunal's operations were suspended as a result of backlash on the part of Zimbabwe. Twenty ICs have never become operational or became defunct after deciding only a few cases (Romano 2014).

ICs have made varying contributions to global governance. First, ICs often contribute to state compliance with international legal commitments, though at times they have sparked defiance, as in the case of the SADC Tribunal. Some ICs have been able to influence the behavior of states, encouraging them to adjust domestic laws to comply with international law. Others provide remedies by identifying appropriate compensatory measures for noncompliance. Second, some ICs have influenced the establishment and the work of other ICs, while others have not had such effects. The International Criminal Court (ICC), for instance, has been shaped by the International Criminal Tribunal for the former Yugoslavia (ICTY) and the International Criminal Tribunal for Rwanda (ICTR). Third, ICs have contributed to broader processes of global governance. The CJEU, for example, has been central to the institutionalization of regional integration in Europe (Burley and Mattli 1993; Mattli and Slaughter 1998; Stone Sweet 2004); the Dispute Settlement Mechanism (DSM) of the World Trade Organization (WTO) has facilitated trade relations globally (Rosendorff 2005; Goldstein et al. 2007; Davis 2012). Evidence has established a link between international criminal prosecutions

and processes of democratization (Sikkink 2011). Some attribute the judicialization of international affairs and constitutionalization of international law to the experiences of the CJEU, ECtHR, the WTO DSM, and others.

As a matter of fact, then, there is a great deal of variation in the *performance* of international courts. This observation raises several critical questions. How should we think about performance in this context? How can we measure performance in specific cases? Why do some ICs perform better than others? What are the determinants of the performance of these courts? Are there ways to improve the performance of international courts? These are the concerns that motivate the contributions to this book.

The Study of Judicial Performance

While prior studies have tended to focus on specific questions relating either to the design or to the effects of international courts, we develop an integrated framework for the study of the performance of ICs. Using this framework, the contributors to this volume present empirical assessments of the performance of international courts that consider both the results that courts produce and the procedures guiding their operation. We explore factors that may explain the patterns of performance we observe.

This study of IC performance takes a broad comparative approach covering the full array of international courts and tribunals. We define the universe of cases to encompass international judicial bodies that: (1) decide the question(s) brought before them on the basis of international law; (2) follow pre-determined rules of procedure; (3) issue legally binding decisions or opinions; (4) are composed of independent members; and (5) require that at least one party to a dispute be a state or an international organization (Romano et al. 2014: 6). While some analysts may argue that international criminal tribunals do not fulfill the fifth criterion, the office of a prosecutor, an organ of an international organization, is one of the parties to disputes in international criminal tribunals (Romano et al. 2014: 7). While ICs can issue nonbinding opinions (i.e., advisory opinions) in addition to legally binding decisions, this definition excludes quasi-judicial bodies, such as the United Nations human rights treaty bodies, which issue only nonbinding recommendations. In addition to permanent bodies, we include ad hoc judicial bodies that meet these

criteria, such as investor-state arbitration tribunals.[4] This allows us to assess the extent to which permanence contributes to performance. We include a range of ICs spanning several issue areas, including human rights, trade, investment, and criminal law.

We are interested both in the outcomes courts produce and the processes through which they arrive at judgments. Drawing on research dealing with the performance of international organizations (Gutner and Thompson 2010), we label these dimensions *outcome performance* and *process performance*. Outcome performance refers to the degree to which ICs attain substantive goals. Process performance, on the other hand, is a matter of the degree to which IC practices measure up to intended or aspired procedural standards.

There are several reasons to study the performance of international courts. For one thing, the performance of these courts affects their legitimacy. As Buchanan and Keohane argue, "If an institution exhibits a pattern of egregious disparity between its actual performance, on the one hand, and its self-proclaimed procedures or major goals, on the other, its legitimacy is seriously called into question" (2006: 422). A court's performance may also affect levels of popular support or the politicization of international legal processes (e.g., see Helfer and Alter 2013). Understanding the performance of ICs has implications as well for the design and reform of international courts. Knowing the determinants of good and bad performance can help in identifying what reform efforts are merited and how reform should proceed. Observers of the European Court of Human Rights, for example, have noted that reform efforts have been informed by policy-relevant assessments of the Court's performance. In addition, the performance of ICs can have implications for the effectiveness of international regimes or governance systems in which they are embedded. Thus, studying IC performance may advance our understanding of effectiveness of international regimes or governance systems.

The study of performance draws on existing research on the performance of international organizations (Gutner and Thompson 2010; Tallberg et al. 2016) and international environmental institutions (Mitchell 2008) as well as research on the effectiveness of international courts (Shany 2014) and international regimes or governance systems (Young 1999; Miles et al. 2002; Hovi et al. 2003; Breitmeier et al. 2006). The literature

[4] Permanence does not refer to whether the court or tribunal itself is permanent. Instead, it means that "they are made of a group of judges who are sitting permanently and are not selected ad hoc by the parties for any given case" (Romano 2011: 262).

on international regimes, governance systems, and social institutions more broadly has focused on the extent to which these arrangements contribute to solving problems (e.g., the depletion of stratospheric ozone, the occurrence of genocide) or steering systems toward socially desirable outcomes (e.g., increased trade, enhanced respect for human rights). The term effectiveness in this literature refers to a measure of the success of regimes and governance systems in solving problems or moving systems toward desired outcomes.

International courts, in our view, are tools or mechanisms that play a number of roles in regimes or governance systems. It is worth differentiating among several distinct roles in this regard. First, courts play roles within individual regimes or governance systems by providing authoritative interpretations of the meaning of a regime's rules, adapting rules to new circumstances, resolving disagreements among parties regarding the meaning of the rules, and helping to ensure compliance on the part of those subject to a regime's rules. Second, courts help to sort out tensions or conflicts between individual regimes or governance systems. This is typically a matter of determining the spheres of applicability of different rules or determining which rules take precedence over others in cases of conflict. Third, courts play a broader role in ensuring that the operations of international regimes or governance systems conform to overarching principles, norms, and values applicable above the level of individual regimes or governance systems. In many cases, this is a matter of ensuring that individual regimes or governance systems adhere to procedural norms like those associated with the idea of fairness.

To know whether regimes or governance systems are effective, then, we argue that it is imperative to ask whether international courts perform their roles well or poorly. The existence of courts that perform well can and often will contribute to the effectiveness of international regimes or governance systems, but there is nothing automatic about this relationship. It is possible to imagine regimes that are successful in solving various problems even in the absence of courts that perform well. Conversely, we can imagine regimes that fail to solve problems, despite the existence of courts that perform their roles well. The relationship between the effectiveness of international regimes or governance systems and the performance of international courts is a matter that merits careful analysis.

Assessing the performance of international courts, on this account, is a matter that extends beyond the intentions of those who create them. Creators may or may not articulate the roles they expect international

courts to play. Moreover, creators may not agree on what roles courts are intended to have. Whether or not the creators are explicit about these matters, however, we can ask questions about the performance of a court in interpreting and adapting rules, sorting out tensions among different regimes, and encouraging adherence to broader principles and norms. As Mitchell observes in his general account of institutional performance, "performance analysis seeks to identify how much an institution contributed to whatever progress was made toward a specified goal" (2008: 79). But as he goes on to say, the relevant goals may be specified by "the creators of the institution, other interested parties, or the evaluator" rather than reflecting only the intentions of the creators (79). In this sense, our focus on performance is similar to Shany's analysis of the "effectiveness" of international courts comparing ". . . actual impacts with desired outcomes, or performance with expectations . . . in the eyes of multiple constituencies" (2014: 6). Unlike Shany, however, we employ the concept of performance to capture a broad set of goals identified by analysts as well as relevant constituencies.

There is an inescapable normative dimension to any study of the performance of international courts. A study of the performance of ICs that looks beyond the intentions of the creators is compatible with evidence suggesting that the effects of these courts go well beyond initial intentions. As Caron argues: "When assessing the value . . . of international courts and tribunals scholars should not only proceed in terms of how well a given institution serves its constituted ends, but also how well it serves the unstated purposes" (2006: 410). More generally, there are normatively grounded differences among observers regarding the proper interpretation of a regime's rules, the precedence granted to one regime vis-à-vis another, and the broader principles to be applied to the operations of regimes. Those who agree that a court performs well, therefore, may be more or less satisfied with the results in normative terms. But this should not detract from efforts to determine how well international courts play their roles and to identify factors that explain variance in performance.

Assessing Performance

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 fulfill 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
>
> (Gutner and Thompson 2010: 231).

We build on this formulation, adopting Gutner and Thompson's distinction between outcomes and processes and adapting it for application to ICs. Thus, one dimension of performance is *outcome performance*, covering the full range of outcomes that result, either directly or indirectly, from the operation of a court.[5] The second dimension is *process performance* associated with the way in which international courts exercise their authority.

To assess both dimensions of performance, we need criteria against which outcomes and processes are compared. As Gutner and Thompson observe, "[e]stablishing 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" (2010: 240). But the selection of appropriate criteria against which performance is compared depends not only on the mandate of a court but also on the perspective of the researcher. Thus, Mitchell notes that while "[i]nstitutions can be evaluated against either the primary or the subsidiary goals for which they were designed . . . they can also be evaluated against the goals of actors outside an institution in question" (2008: 80).

It follows that the criteria of evaluation can vary considerably. Still, we endeavor to develop realistic criteria based on common understandings of international courts. We start by identifying the functions or roles that analysts typically ascribe to international courts. Some of these are clearly intended by creators of ICs, as reflected in their constitutive documents. Others are roles that ICs serve in practice, irrespective of creators' intentions. We then posit that these are plausible criteria against which the performance of courts can be evaluated. We do not claim that these criteria apply equally to all courts. Nor do we assume that they are exhaustive. Rather, our intention is to provide a common analytical starting point for comparison across ICs. We recognize that it is necessary to contextualize

[5] We use outcome here as a generic term to encompass a broad set of results. For these purposes, we do not distinguish outcomes from outputs and impacts, as is typical in comparative politics literature or in regime effectiveness literature. In other words, the outputs of an international court (or their judgments) are among the range of outcomes resulting from the IC's operation. The same applies to impacts.

criteria on a case-by-case basis, which all the contributions to this volume do. We also recognize that courts rarely bear sole responsibility for the production of specific outcomes. Our objective is to make judgments regarding the proportion of the variance in outcomes that we can attribute to the activities of specific courts.

The analysis of *outcome performance* is a matter of assessing the results of court activities. This means evaluating the extent to which the actions of courts contribute to the attainment of substantive goals. The relevant goals may be framed at the level of specific cases, the level of the governance system to which cases belong, or the level of the development of international society as a whole. Thus, the relevant criteria may range from a narrow conception of performance to much broader conceptions. Drawing on existing literature, however, we approach outcome performance in terms of the roles or functions of courts in various social settings. Specifically, we identify criteria dealing with dispute settlement, clarification of the law, and compliance.

The first criterion involves the function of settling disputes. To what extent are courts able to settle disputes? How effectively do they do so? Extensive research suggests that international courts and tribunals do contribute to the settlement of disputes (Alter 2013; Bogdandy and Venzke 2013; Alvarez 2014; Shany 2014). Traditional perspectives on the dispute settlement function of ICs focus on the settlement of disputes between states. More recently, several ICs have begun to address disputes between private parties as well as states. Human rights courts, for instance, often have jurisdiction over disputes between private actors and states. International organizations also may be involved in cases that come before ICs, as in cases where the UN General Assembly requests advisory opinions from the ICJ, or when the CJEU reviews disputes between the Commission of the European Union and a member state. While dispute settlement is generally recognized as an important role of ICs, it is not a criterion that applies to all courts. Many scholars conclude that international criminal courts or tribunals prosecute crimes rather than acting to settle disputes (Alter 2013; Bogdandy and Venzke 2013). Still, dispute settlement is an important focus in our examination of the performance of ICs. Are they good at settling disputes? What enables courts to settle disputes successfully?

A second criterion for assessing the outcome performance of international courts is clarification of the law. Most treaties establishing courts assign them the task of "interpreting" the law. Several scholars also emphasize the importance of clarifying the law (Alter and Helfer 2010;