



Governance by Indicators

Global Power through Quantification and Rankings

Edited by Kevin E. Davis, Angelina Fisher,
Benedict Kingsbury, and Sally Engle Merry

LAW AND GLOBAL GOVERNANCE

OXFORD

GOVERNANCE BY INDICATORS

LAW AND GLOBAL GOVERNANCE SERIES

Editors:

Andrew Hurrell, Benedict Kingsbury, and Richard B. Stewart

Global governance involves the exercise of power, beyond a single state, to influence behaviour, to generate resources, or to allocate authority. Regulatory structures, and law of all kinds, increasingly shape the nature, use, and effects of such power. These dynamic processes of ordering and governance blend the extra-national with the national, the public with the private, the political and economic with the social and cultural. Issues of effectiveness, justice, voice, and inequality in these processes are growing in importance. This series features exceptional works of original research and theory—both sector-specific and conceptual—that carry forward the serious understanding and evaluation of these processes of global governance and the role of law and institutions within them. Contributions from all disciplines are welcomed. The series aims especially to deepen scholarship and thinking in international law, international politics, comparative law and politics, and public and private global regulation. A major goal is to study governance globally, and to enrich the literature on law and the nature and effects of global governance beyond the North Atlantic region.

Governance by Indicators

Global Power through Quantification and Rankings

Edited by

KEVIN E. DAVIS, ANGELINA FISHER,
BENEDICT KINGSBURY,
and
SALLY ENGLE MERRY



INSTITUTE FOR INTERNATIONAL LAW AND JUSTICE
NEW YORK UNIVERSITY SCHOOL OF LAW

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© The several contributors 2012

The moral rights of the authors have been asserted

First Edition published in 2012

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Crown copyright material is reproduced under Class Licence
Number C01P0000148 with the permission of OPSI
and the Queen's Printer for Scotland

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloguing in Publication Data
Library of Congress Control Number: 2012936927

ISBN 978-0-19-965824-4

Printed in Great Britain
on acid-free paper by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Acknowledgements

This book is one result of an ongoing project on indicators and governance by information, conducted by the Institute for International Law and Justice at New York University School of Law. The project is closely connected with the IILJ's major project on global regulatory governance and global administrative law. Further materials of these projects are at iilj.org. This is also the first volume in the Law and Global Governance series published by Oxford University Press.

The editors are very grateful to Carnegie Corporation of New York, the Rockefeller Foundation, and the National Science Foundation for grants to the IILJ which have made this work possible. Many valuable comments on the project and on different chapters were received from (among others) Juan Botero, Elizabeth Boyle, Christopher Bradley, Bruce Carruthers, Ariel Colonosmos, Alex Cooley, Grainne de Burca, Megan Donaldson, Marion Fourcade, Tom Ginsburg, Ryan Goodman, Andrew Hurrell, Robert Keohane, Andrew Lang, Emily Martin, Amanda Perry-Kessaris, Mariana Mota Prado, Richard Rottenburg, Charles Sabel, Galit Sarfaty, Greg Shaffer, Jack Snyder, Richard Stewart, Kay Warren, and Joseph Weiler, and from Zia Khan, Claudia Juech, Evan Michelson and several of their colleagues at presentations at the Rockefeller Foundation, the participants in the ongoing NSF-funded network on indicators, participants in the projects of the Global Administrative Law Network very generously supported by the International Development Research Centre of Canada, and the contributors to this volume. Annual meetings of the Law and Society Association (IRC 25) and the American Society of International Law have enabled very helpful collective discussions of this work. We thank Dean Richard Revesz of NYU Law School, and Merel Alstein, Alex Flach and Briony Ryles of Oxford University Press, for their support and encouragement, and Rachel Jones, Nikki Reisch, Kaveri Vaid and Shannon Xydis in the IILJ for production assistance. We thank also Chris Jordan for kindly allowing use of his art on the cover, and Lorenzo Casini for drawing our attention to Chris Jordan's *Running the Numbers* series.

KED, AF, BK, and SEM

Table of Contents

Abbreviations
List of Contributors

viii
xii

PART I: INDICATORS AS TECHNOLOGIES OF KNOWLEDGE PRODUCTION AND GLOBAL GOVERNANCE: KEY CONCEPTS AND APPROACHES

1. Introduction: Global Governance by Indicators <i>Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry</i>	3
2. Beyond Supply and Demand: A <i>Political</i> -Economic Conceptual Model <i>Tim Büthe</i>	29
3. Taming and Framing Indicators: A Legal Reconstruction of the OECD's Programme for International Student Assessment (PISA) <i>Armin von Bogdandy and Matthias Goldmann</i>	52
4. The Dynamism of Indicators <i>Wendy Nelson Espeland and Michael Sauder</i>	86
5. Semiotics of Indicators: The Case of Corporate Human Rights Responsibility <i>Ronen Shamir and Dana Weiss</i>	110
6. Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order <i>Nehal Bhuta</i>	132

PART II: INDICATORS, POWER, AND AUTHORITY IN GLOBAL GOVERNANCE

7. Re-construction of Private Indicators for Public Purposes <i>Katharina Pistor</i>	165
8. Legal Yardsticks: International Financial Institutions as Diagnosticians and Designers of the Laws of Nations <i>Terence C. Halliday</i>	180
9. From Diagnosing Under-immunization to Evaluating Health Care Systems: Immunization Coverage Indicators as a Technology of Global Governance <i>Angelina Fisher</i>	217

PART III: TRANSLATION, TRANSPLANTATION,
AND ADAPTATION: THE RELATIONSHIP
BETWEEN “GLOBAL” AND “LOCAL” IN
INDICATOR PRODUCTION AND USE

10. Internally Displaced Population in Colombia: A Case Study on the Domestic Aspects of Indicators as Technologies of Global Governance 249
René Urueña
11. Problems of Power in the Design of Indicators of Safety and Justice in the Global South 281
Christopher Stone

PART IV: CASE STUDIES: ASSESSING THE
STRENGTHS, PROBLEMS, AND EFFECTS OF
INDICATORS IN HUMAN RIGHTS, HUMANITARIAN
ASSISTANCE, AND SOCIAL INVESTMENT

12. Measuring Human Rights: UN Indicators in Critical Perspective 297
AnnJanette Rosga and Margaret L. Satterthwaite
13. The Use of Indicators to Measure Government Responses to Human Trafficking 317
Anne T. Gallagher and Janie Chuang
14. Fighting Human Trafficking or Instituting Authoritarian Control? The Political Co-optation of Human Rights Protection in Belarus 344
Marina Zoloznaya and John Hagan
15. Rights-based Humanitarian Indicators in Post-earthquake Haiti 365
Margaret L. Satterthwaite
16. Impact Investment Indicators: A Critical Assessment 392
Sarah Dadush

PART V: REGULATING INDICATORS

17. Accountability in the Generation of Governance Indicators 437
Nikhil K. Dutta
18. Public Regulation of Global Indicators 465
Sabino Cassese and Lorenzo Casini

- Index* 475

Abbreviations

ACER	Australian Council for Educational Research
ADB	Asian Development Bank
ALNAP	Active Learning Network for Accountability and Performance
AMC	Advanced Market Commitment
ARWU	Shanghai Jiao Tong Academic Ranking of World Universities
AusAID	Australian Agency for International Development
AUSAID	Australian Agency for International Development
BERI	Business Environment Risk Intelligence
BI	Business Intelligence
BMZ	Federal Ministry for Economic Cooperation and Development (Germany)
BPC	Board of Participating Countries
BYU	Brigham Young University
CCCM	Camp Coordination and Camp Management
CERI	Centre for Educational Research and Innovation
CESCR	Committee on ESCR
CESR	Committee of European Securities Regulators
CIA	Central Intelligence Agency
CIDA	Canada's International Development Agency
CIR	Camp Indicator Report
CIS	Commonwealth
CMA	camp management agencies
CODHES	Consulatoria para los Derechos Humanos y el Desplazamiento
COP	Conference of the Parties
COP	Communication on Progress (chapter 5)
CPIA	Country Policy and Institutional Assessment
DESA	Department of Economic and Social Affairs
DFID	British Department for International Development
DHO	District Health Office
DIHR	Danish Institute for Human Rights
DOT	directly observed treatment
DQA	data quality audit
DQS	data quality self-assessment
DTP3	third dose of diphtheria-tetanus-pertussis vaccine
EBRD	European Bank for Reconstruction and Development
EFE	A Colombia News Channel
EPI	Expanded Programme on Immunization
ESMA	European Securities and Markets Authority

FCO	Foreign and Commonwealth Office
FDI	foreign direct investment
FH	Freedom House
GAL	Global Administrative Law
GAVI	Global Alliance for Vaccines and Immunization
GBV	gender-based violence
GC	Global Compact
GHO	Global Health Observatory
GIIRS	Global Impact Investment Rating System
GPA	Grade Point Average
GRETA	Group of Experts on Action against Trafficking in Human Beings
GRI	Global Reporting Initiative
GTIP	United States Government Office to Monitor and Combat Trafficking in Persons
GTZ	Gesellschaft für Technische Zusammenarbeit
HAP	Humanitarian Accountability Partnership
HDI	Human Development Index
HIPC	Heavily Indebted Poor Countries
HNP	Health, Nutrition and Population Family
IASC	Inter-Agency Standing Committee
ICRC	International Committee for the Red Cross
ICRG	International Country Risk Guide
IDA	International Development Association
IDMC	Internal Displacement Monitoring Centre
IDP	international development programme
IEA	International Association for the Evaluation of Educational Achievement
IEC	International Electrotechnical Commission
IESCR	International Covenant on Economic, Social and Cultural Rights
IFC	International Finance Corporation
IFIs	inter-governmental financial institutions
IFIs	international financial institutions
IFRC	International Federation of the Red Cross and Red Crescent Societies
IHME	Institute for Health Metrics and Evaluation
IMF	International Monetary Fund
INES	Indicators of Education Systems
INGOs	international NGOs
IO	international organization
IOSCO	International Organization of Securities Commissions
IRIS	Impact Reporting and Investment Standards
ISS	Immunisation Services Support
ITLOs	international trade law organizations
JRP	Joint Reporting Form on Immunization
LSAT	Law School Admission Test

MAR	Minorities at Risk
MCC	Millennium Challenge Corporation
MCV1	Molluscum contagiosum virus type 1
MIOB	Making It Our Business (framework)
MOI	Ministry of the Interior
MYP	multi-year plan
NGO	non-governmental organization
NPA	national policy assessment
NYU	New York University
OCHA	Office for the Coordination of Humanitarian Affairs
OECD	Organization for Economic Co-operation and Development
OHCHR	UN Office of the High Commissioner for Human Rights
PGB	PISA Governing Board
PHA Indicators	Public Health Activity Performance Indicators
PHC	primary health care
PISA	Programme for International Student Assessment
PRS	Political Risk Services
PRSP	Poverty Reduction Strategy Paper
QJE	Quarterly Journal of Economics
QSWUR	QS World University Rankings
RBM	results-based management
ROSC	Report on Observance of Standards and Codes
RPNGC	Royal Papua New Guinea Constabulary
SEC	US Security and Exchange Commission
SI	système international (international system of units)
SIR	Standards and Indicators Report
SISDHES	Information System on Forced Displacement and Human Rights (initials in Spanish)
SMEs	small and medium-sized businesses
TIMSS	Trends in International Mathematics and Science Study
TVPA	Trafficking Victims Protection Act 2000
TVPRA	Trafficking Victims Protection Reauthorization Act 2005
UN	United Nations
UNCITRAL	UN Commission on International Trade Law
UNDP	UN Development Program
UNESCO	UN Educational, Scientific and Cultural Organization
UNFPA	UN Population Fund
UNGA	UN General Assembly
UNHCR	UN High Commissioner for Refugees
UNICEF	UN Children's Fund
UNIDROIT	International Institute for the Unification of Private Law

UNM	University of New Mexico
UNRRA	United Nations Relief and Reconstruction Agency
UNSD	United Nations Statistics Division
USAID	US Agency for International Development
WASH	water, sanitation, and hygiene
WB	World Bank
WGs	Worldwide Governance Indicators
WHO	World Health Organization

List of Contributors

Nehal Bhuta, European University Institute, Florence.

Tim Büthe, Department of Political Science, Duke University.

Lorenzo Casini, University of Rome La Sapienza.

Sabino Cassese, Italian Constitutional Court.

Janie Chuang, American University Washington College of Law.

Sarah Dadush, International Fund for Agricultural Development, Rome.

Kevin E. Davis, New York University School of Law.

Nikhil K. Dutta, Law Clerk, U.S. Court of Appeals, Boston.

Wendy Nelson Espeland, Northwestern University.

Angelina Fisher, New York University School of Law.

Anne T. Gallagher, Independent.

Matthias Goldmann, Max Planck Institute for Comparative Public Law and International Law.

John Hagan, Northwestern University and American Bar Foundation.

Terence C. Halliday, American Bar Foundation.

Benedict Kingsbury, New York University School of Law, Visiting Professor, University of Utah.

Sally Engle Merry, New York University School of Law.

Katharina Pistor, Columbia Law School.

AnnJanette Rosga, Women's International League for Peace and Freedom.

Margaret L. Satterthwaite, NYU School of Law.

Michael Sauder, University of Iowa.

Ronen Shamir, Tel Aviv University.

Christopher Stone, Kennedy School of Government, Harvard University; President, Open Society Foundations (as of July 1, 2012).

René Uruña, Los Andes University.

Armin von Bogdandy, Max Planck Institute for Comparative Public Law and International Law.

Dana Weiss, Tel Aviv University.

Marina Zaliznaya, Northwestern University.

PART I

INDICATORS AS TECHNOLOGIES OF KNOWLEDGE PRODUCTION AND GLOBAL GOVERNANCE: KEY CONCEPTS AND APPROACHES

This page intentionally left blank

1

Introduction: Global Governance by Indicators

*Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry**

I. Introduction

The production and use of indicators in global governance is increasing rapidly. Users include public international development agencies such as the World Bank and the United Nations, national governmental aid agencies such as the US government's Millennium Challenge Corporation (MCC), global businesses and investors; bodies concerned with assessing or enforcing compliance with existing legal standards such as human rights treaty monitoring bodies, advocacy groups including many NGOs, and various scientific or expert communities, especially in the field of political science. Examples of prominent indicators and their producers or promulgators include: Doing Business Indicators produced by the International Finance Corporation (a member of the World Bank Group); Governance Indicators, including The Control of Corruption and Rule of Law, under the imprimatur of the World Bank; the Millennium Development Goals indicators under UN auspices; the Corruption Perceptions Index created by Transparency International; the Human Development Index (HDI) produced by the United Nations Development Program (UNDP); the Trafficking in Persons indicators produced by the US State Department; and various indicators produced by consultancies specializing in advising investors on political risks. The Office of the United Nations High Commissioner for Human Rights has explored possibilities of developing indicators for several core human rights.

* The authors acknowledge with thanks National Science Foundation grants #SES-0921368 and #SES-023717, the latter to support a multi-country collaborative project focused on indicators in developing and transitional countries. The first part of this chapter draws extensively, with permission, on Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance," *Law and Society Review* 46 (2012). Many helpful suggestions and contributions were made by Angelina Fisher. The authors are grateful also for research assistance by Maxwell Kardon and Jessica Shimmin, and for generous financial support for related work from Carnegie Corporation of New York, the Straus Institute for the Advanced Study of Law and Justice at NYU, NYU Law School's D'Agostino and Greenberg Faculty Research Fund, and the Rockefeller Foundation.

The burgeoning production and use of indicators in global governance has the potential to alter the forms, the exercise, and perhaps even the distributions of power in certain spheres of global governance. Yet the increasing use of indicators has not been accompanied by systematic study of and reflection on the implications, possibilities and pitfalls of this practice. As a result, little attention has been paid to questions such as: "What social processes surround the creation and use of indicators?," "How do the conditions of production influence the kinds of knowledge that indicators provide?," "How does the use of indicators in global governance change the nature of standard-setting and decision-making?," "How does it affect the distribution of power between and among those who govern and those who are governed?," "What is the nature of responses to the exercises of power through indicators, including forms of contestation and attempts to regulate the production or use of indicators?" The answers to these questions all have significant normative, theoretical and practical implications.

This book is part of a project on *Indicators as a Technology of Global Governance* of the Institute for International Law and Justice at New York University School of Law, which seeks to explain the phenomena of indicators, ranking and measurements in global governance and to understand their impact on countries and institutions being evaluated. The project is a multi-disciplinary enterprise drawing on perspectives from political science, sociology, anthropology, and law, and comprising both theoretical inquiries and case studies. In this book, prominent sociologists, anthropologists, political scientists and legal scholars use a power-knowledge framework to study the effects of quantification and indicators on decision-making, resource allocation, social categories, forms of contestation and power of experts within and across institutions. Each chapter examines indicators in a particular sector of practice while presenting arguments and insights of wider significance. In doing so, these investigations draw on several existing bodies of scholarship. Work in three areas may be highlighted.

First, several contributors use portions of the substantial body of work on connections of law and power in global governance.¹ This includes scholarship

¹ Balakrishnan Rajagopal, *International Law from Below: Development, Social Movements, and Third World Resistance* (Cambridge: Cambridge University Press, 2003); John Braithwaite, "Methods of Power for Development: Weapons of the Weak, Weapons of the Strong," *Michigan Journal of International Law* 26 (2004): 298–330; Anne-Marie Slaughter, *A New World Order* (Princeton, NJ: Princeton University Press, 2004); Boaventura de Sousa Santos and Cesar A. Rodriguez-Garavito (eds), *Law and Globalization from Below: Towards a Cosmopolitan Legality* (Cambridge, UK: Cambridge University Press, 2005); Benedict Kingsbury et al., "The Emergence of Global Administrative Law," *Law and Contemporary Problems*, 68 (Summer/Autumn 2005): 15–61; Sally Engle Merry, *Human Rights and Gender Violence: Translating International Law into Local Justice* (Chicago: University of Chicago Press, 2006); Mark Goodale and Sally Engle Merry (eds), *The Practice of Human Rights: Tracking Law Between the Global and the Local* (Cambridge: Cambridge University Press 2007); Terence C. Halliday and Bruce G. Carruthers, *Bankrupt: Global Lawmaking and Systemic Financial Crisis* (Stanford, CA: Stanford University Press, 2009); Benedict Kingsbury, "The Concept of 'Law' in Global Administrative Law," *European Journal of International Law* 20 (2009): 23–57; Beth A. Simmons, *Mobilizing for Human Rights: International Law in Domestic Politics* (Cambridge, UK: Cambridge University Press, 2009).

dealing with “new governance” and experimentalist learning models,² with theories of governmentality,³ and with networks.⁴

A second starting point is theoretical writings on quantification and indicators as social phenomena, both general works⁵ and a small but growing body of studies relating to specific uses of indicators and quantification in global governance contexts.⁶

Third, important insights and perspectives on indicators come from science and technology studies (STS),⁷ including actor network theory.⁸

² Gráinne De Búrca, “New Governance and Experimentalism: An Introduction,” *Wisconsin Law Review* 2 (2010): 227–39; “Symposium: New Governance and the Transformation of Law,” *Wisconsin Law Review* (2010): 227–748. Charles Sabel and Jonathan Zeitlin (eds), *Experimentalist Governance in the European Union: Towards a New Architecture* (Oxford: Oxford University Press, 2010).

³ Peter Miller and Nikolas Rose, *Governing the Present* (Cambridge: Polity Press, 2008).

⁴ Bruno Latour, “Préface: Le fantôme de l’esprit public—Des illusions de la démocratie aux réalités de ses apparitions,” in Walter Lippmann, *Le public fantôme* (Paris: Editions Demopolis, 2008); Bruno Latour, “Networks, Societies, Spheres: Reflections of an Actor-Network Theorist,” *International Journal of Communication* 5 (2011): 796–810.

⁵ Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990); Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press 1995); Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, MA: Harvard University Press, 1998); Wendy Nelson Espeland and Mitchell L. Stevens, “A Sociology of Quantification,” *European Journal of Sociology* 49 (2008): 401–36; Wendy Nelson Espeland and Michael Sauder, “Rankings and Reactivity: How Public Measures Recreate Social Worlds,” *American Journal of Sociology* 113 (2007): 1–40; Peter Andreas and Kelly Greenhill (eds), *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict* (Ithaca: Cornell University Press, 2010); Ann Rudinow Saetnan et al. (eds), *The Mutual Construction of Statistics and Society* (New York: Routledge, 2011).

⁶ Christiane Arndt and Charles Oman, *Uses and Abuses of Governance Indicators* (Paris: OECD Development Centre Study, 2006); Kevin E. Davis and Michael B. Kruse, “Taking the Measure of Law: The Case of the Doing Business Project,” *Law & Social Inquiry* 32 (2007): 1095–119; Christopher Hood et al., “Rating the Rankings: Assessing International Rankings of Public Service Performance,” *International Public Management Journal* 11 (2008): 298–358; Christiane Arndt, “The Politics of Governance Ratings,” *International Public Management Journal* 11 (2008): 275–97; Armin von Bogdandy and Matthias Goldmann, “The Exercise of International Public Authority through National Policy Assessment: The OECD’s PISA Policy as a Paradigm for a New International Standard Instrument,” *International Organizations Law Review* 5 (2008): 241–98; Tore Fougner, “Neoliberal Governance of States: The Role of Competitiveness Indexing and Country Benchmarking,” *Millennium Journal of International Studies* 37 (2008): 303–26; AnnJanette Rosga and Margaret L. Satterthwaite, “The Trust in Indicators: Measuring Human Rights,” *Berkeley Journal of International Law* 29 (2009): 256–315; Martin Ravallion, “Troubling Tradeoffs in the Human Development Index,” World Bank Policy Research Working Paper 5484 (Washington DC: World Bank, 2010); Margaret Satterthwaite, “Indicators in Crisis: Rights-based Humanitarian Indicators in Post-earthquake Haiti,” *New York University Journal of International Law & Politics* 43 (2011): 865–964; Sally Engle Merry, “Measuring the World: Indicators, Human Rights, and Global Governance,” *Current Anthropology* 52 (Supp. 3) (2011): S83–95.

⁷ Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences* (Cambridge, MA: The MIT Press, 1999); Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society* (Cambridge, MA: Harvard University Press, 1987); Martha Lampland and Susan Leigh Star, *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life* (Ithaca: Cornell University Press, 2009); Ann Rudinow Saetnan et al. (eds), *The Mutual Construction of Statistics and Society* (New York: Routledge, 2011).

⁸ Bruno Latour, *Reassembling the Social* (Oxford, UK: Oxford University Press, 2005); Bruno Latour, “Networks, Societies, Spheres: Reflections of an Actor-Network Theorist” (n. 4).

Part II of this chapter sets out our conceptual claims regarding the defining characteristics of indicators. Part III identifies defining features of governance and global governance and sets out several hypotheses concerning the reasons for, and the implications of, the turn to indicators in global governance. Part IV provides an overview of the volume and draws attention to ways in which some of the insights in particular chapters have general significance, including in relation to the hypotheses formulated in Part III. Part V concludes.

II. What is an indicator?

a. Indicators defined

There is no agreed meaning of “indicator,” but for the purposes of our inquiry into indicators as an important emerging technology in the practice of global governance the concept can be delimited in the following way.

An indicator is a named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis (such as countries or institutions or corporations), synchronically or over time, and to evaluate their performance by reference to one or more standards.

This working definition subsumes indexes, rankings, and composites which aggregate different indicators. Many of the best-known indicators are aggregations or “mash-up” compilations,⁹ with substantial discretion available to the compiler in choosing what specific indicators to include, with what weightings and what devices to limit double-counting or to smooth over data unavailability. Examples include the HDI and the World Governance Indicators. While the processes and uses of aggregation raise many special issues, for the purposes of this volume the term “indicators” also includes these aggregations. We focus on the subset of indicators that are used for evaluation or judgment, and have effects specifically on decision-making or other effects in global governance. The term is also used in other ways—for example, to refer to a diagnostic characteristic (such as an indicator of a person who has been trafficked, or an indicator species for an ecosystem)—but these usages are outside the concept as used by the contributors to this volume.

Indicators often take the form of, or can readily be transformed into, numerical data. A key challenge is whether and how indicators ought to be distinguished from other compilations of numerically rendered data. The differences lie in how indicators simplify “raw” data and then name the resulting product. That simplification can involve aggregation of data from multiple sources. It can also involve filtering that excludes certain data, including outliers or other data deemed to be

⁹ Martin Ravallion, “Mashup Indices of Development,” Policy Research Paper No. 5432 (Washington, DC: The World Bank, 2010).

unreliable or irrelevant. Sometimes data are filtered out and replaced with statistics, such as means or standard deviations, meant to convey similar information. In still other cases missing data are filled in with values estimated from existing data. The specific name given to data that have been organized and simplified in these ways typically denotes the social phenomenon the data ought to be taken to represent. So for example, a census report containing data on the numbers of people between the ages of 0–14, 15–64, and 65 + is not in itself an indicator. But suppose that data is aggregated in a particular way, for instance by dividing the sum of the first and third figures by the figure for the number of people in the 15–64 group. If that number is then labeled a “dependency ratio,” and the same calculation is made for other units or other times, the collection of processed data is capable of being used for the purposes of inter-country or inter-temporal comparisons of “dependency” and qualifies as an indicator.

Indicators can also be contrasted with other representations of social phenomena. In principle, any given social phenomenon can be represented in multiple ways. For example, the level of respect for the rule of law in a given country in a given year may be represented by an indicator such as a rule of law index. Alternatively, however, it might be represented by a paragraph of text describing patterns of respect or disregard for the rule of law during the relevant period, or by a series of striking photographs or a video recording. All of these representations may purport to capture the same phenomenon. Each involves some form of simplification (although the forms vary), and each may be given a suggestive name by its producer. However, the indicator is distinctive in the ways in which it represents and conveys compiled numerical data, and it has particular attractions as a means of representation for use in comparing or evaluating particular units of analysis. Different representations are likely to convey different impressions and stimulate different responses, in ways that vary with the type of audience. Indicators cater to demand for (and receptivity to) numerical, rank-ordered and comparable data.

There is considerable room for variation within the scope of our broad definition of an indicator. Some indicators have names that are highly evocative of evaluative standards; some provide more complete orderings of the units being analyzed; some involve greater simplification of raw data. The indicators addressed in different chapters of this volume vary along each of these continua.

b. Salient characteristics of indicators

Our working definition highlights several features of indicators, including (1) the significance of the name of the indicator and the associated assertion of its power to define and represent a phenomenon such as “the rule of law”; (2) the ordinal structure enabling comparison and ranking and exerting pressure for “improvement” as measured by the indicator; (3) the simplification of complex social phenomena; and (4) the potential to be used for evaluative purposes. We elaborate on the significance of these features in the following paragraphs.

1. Naming the indicator

The assertion that an indicator has been brought into existence and given life is typically marked by naming it. The name itself is usually a simplification of what the index purports to measure or rank. The name's constancy may mask changes over time in the indicator itself. Calling an indicator a measure of "transparency" or "human development" asserts a claim that there is such a phenomenon and that the numerical representation measures it. An indicator may even create the phenomenon it claims to measure, as IQ tests came to define intelligence. Labeling this measure an Indicator, Index, Ranking, League Table, etc. implies a claim to knowing and measuring a phenomenon. As a result, the indicator represents an assertion of power to produce knowledge and to define or shape the way the world is understood.

2. Rank-ordered structure

All indicators are fundamentally comparative, and some element of ranking is a feature of the indicators we are studying. Indicators usually enable comparison of different units, but in a few cases only permit comparison of the same unit at different times. However, an indicator need not rank all data points or all units in a transitive way. Influential indicators are usually cardinal (attributing separately defined values to each unit), and most use one or other of a standard menu of scaling methods (e.g., a purely ordinal scale, an equal-interval scale, or a ratio scale), but it is possible to have an indicator which does not have these attributes. Some listings with most of the attributes of indicators may merely divide units into categories described nominally, identifying difference without ranking the categories. These do not fall within our definition of an indicator. Other nominal listings may have an element of hierarchy among broad categories (red, yellow, green). These do qualify as indicators for our purposes.

3. Simplification

Simplification (or reductionism) is central to the appeal (and probably the impact) of indicators. They are often numerical representations of complex phenomena intended to render these simpler and more comparable with other complex phenomena which have also been represented numerically. Indicators are typically aimed at policymakers and are intended to be convenient, easy to understand, and easy to use. Yet, the transformation of particularistic knowledge into numerical representations that are readily comparable strips meaning and context from the phenomenon. In this numerical form, such knowledge carries a distinctive authority that shifts configurations and uses of power and of counter-power. This transformation reflects, but also contributes to, changes in decision-making structures and processes.

Indicators also often present the world in black and white, with few ambiguous intermediate shades. They take flawed and incomplete data that may have been

collected for other purposes, and merge them together to produce an apparently coherent and complete picture. Wendy Espeland and Mitchell Stevens identify this as a potential consequence of what March and Simon refer to as uncertainty absorption, which “takes place when inferences are drawn from a body of evidence, and the inferences instead of the evidence itself, are then communicated.”¹⁰ As Espeland and Stevens describe this process, “Raw” information typically is collected and compiled by workers near the bottom of organizational hierarchies; but as it is manipulated, parsed, and moved upward, it is transformed so as to make it accessible and amenable for those near the top, who make the big decision. This “editing” removes assumptions, discretion and ambiguity, a process that results in “uncertainty absorption”: information appears more robust than it actually is... the premises behind the numbers disappear, with the consequence that decisions seem more obvious than they might otherwise have been. An often unintended effect of this phenomenon is numbers that appear more authoritative as they move up a chain of command. The authority of the information parallels the authority of its handlers in the hierarchy.¹¹

The degree of uncertainty beneath the surface of many of the most influential simplifying indicators in global governance is quite intensively scrutinized, but usually only in specialized scientific literature.¹²

4. Indicators as tools for evaluation

We single out indicators from other collections of data based on their potential use in evaluating performance. Indicators set standards. The standard against which performance is to be measured is often suggested by the name of the indicator—corruption, protection of human rights, respect for the rule of law, etc. To the extent that an indicator is used to evaluate performance against one standard rather than another, the use of that indicator embodies a theoretical claim about the appropriate standards for evaluating actors’ conduct. Indicators often have embedded within them, or are placeholders for, a much more far-reaching theory—which some might call an “ideology”—of what a good society is, or how governance should ideally be conducted to achieve the best possible approximation of a good society or good policy. At a minimum they are produced as, or used as, markers for larger policy ideas. They may measure “success” directly along this axis, or they may measure what, from the standpoint of the theory or policy idea, are pathologies or

¹⁰ James G. March and Herbert A. Simon, *Organizations* (New York: Wiley, 1958), 165.

¹¹ Espeland and Stevens, “A Sociology of Quantification” (n. 5), 421–2.

¹² Stephen Morse, *Indices and Indicators in Development: An Unhealthy Obsession with Numbers?* (London: Earthscan, 2004); Hood et al., “Rating the Rankings: Assessing International Rankings of Public Service Performance” (n. 6); Bjorn Hoyland et al., “The Tyranny of International Index Rankings,” *Journal of Development Economics* 97 (2012): 1–14. Hoyland, Moene, and Willumsen reintroduce the uncertainty that is filtered out of the Human Development Index indicators and the Doing Business indicators when these are aggregated into rankings. They calculate, for example, that the ranking of the top four countries as shown in the 2008 HDI has less than 1 percent probability of being the true top rank. The estimated confidence intervals for Georgia in the 2007 Doing Business rankings are 11th place to 59th place, rather than the definitive 18th place Georgia was given in the report.

problems to be overcome. More frequently they address simply some measurable elements within a wider scenario envisaged by the theory or policy idea. Often the theory or policy idea is not spelled out at all in the indicator, but remains implicit.¹³

The theory or idea embedded in an indicator may be developed or reframed by its users or by other actors in ways that differ from anything intended by the producers. Indicators often express ideologies about the ideal society and the process of achieving it. But what they actually communicate, and to whom, may not be what their producers and promulgators sought to communicate. This communicative element makes it essential to consider the indicator's audience and how it is engaged by the indicator.

Use of the indicator in evaluative processes requires that its audience include active evaluators. Those evaluators may or may not exert significant governance power over the actor being evaluated. An indicator may be taken up by its audience (sometimes without any explicit intention on their part) in social processes that do not directly involve evaluation, including: establishing or cementing key concepts (such as "human development"); influencing actor identities; condensing and redefining status and hierarchies in quantified forms, framing standards or causal theories which may then be rendered in other ways (for example, in an organizational policy or a statement of best practices); influencing decisions as to what is measured or how statistics are compiled; crudely validating and calling into question other ideas or evaluative impressions. These other roles or uses of indicators do not alter the definitional requirement that an indicator must be *capable* of being used for evaluation, even while some of its roles and effects do not depend on the operation of specific evaluative processes.

III. Indicators as technologies of global governance

a. Global governance defined

Governance comprises the means used to influence behavior, the production of resources and the distribution of resources. Thus governance is a broader concept than regulation, which refers to means used to influence the behavior of regulated actors (the *Regulation and Governance* approach); however, the distinction is often a fine one because the process of allocating resources, and even the process of generating or not generating resources, can also serve as a means of regulation. Analyses of the means and impacts of governance vary in focus. Some address mainly material allocations and influences, as in the epigram that politics is who gets what, when, and how.¹⁴ Others in Foucauldian or Marxian veins are concerned with the impact of power relations on identity and consciousness, the

¹³ Poovey suggests that the use of numerical information to understand the world in ways that appear objective and free from interpretation but obscure underlying theoretical assumptions is a distinctive feature of modernity. Mary Poovey, *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society* (Chicago, IL: University of Chicago Press, 1998).

¹⁴ Harold D. Lasswell, *Politics: Who Gets What, When, How* (Cleveland: Meridian Books, 1936).

constitution of the subject, and the analysis of structures of power or domination which the actors may not themselves be aware of. Others examine governance in the interactions of largely autonomous systems,¹⁵ or in self-organizing systems that lack apparent intentionality,¹⁶ or in certain actor-network forms that have not (or not yet) supported the delineation and articulation of forms of authority and governance.¹⁷

In many situations across this range (but not all), governance can be modeled using a standard triangular schematic which posits relations between the actors (the *governors*) who allocate resources among or exert influence over the behavior of other actors, the actors subject to governance (the *governed*), and other interested constituencies (the *public*).¹⁸

The process of governance is often itself subject to governance. In other words, governors are often simultaneously among the governed, in the sense that their actions are typically subject to various forms of contestation and control. Contestation can take many forms, including, depending on the context, violence, deliberate non-compliance, litigation, behind-the-scenes lobbying, or voting. Meanwhile control can range from resistance to specific decisions concerning specific actors, to much more systematic and generalized efforts at regulation. This last scenario can involve what Grabosky describes as “layers of regulation,” citing situations in which private actors who serve as regulators are in turn subject to monitoring and control by public actors (see Figure 1.1).¹⁹

Governance can be effected through a wide variety of mechanisms, including military action, transfers of funds, promulgation of legal instruments, publication of scientific reports, advertising campaigns, or educational programs. Following Miller and Rose,²⁰ we call such mechanisms “technologies” of governance.²¹ Different technologies of governance involve generation and allocation of different kinds of resources, including both material resources such as money or personnel, and intangible resources such as status and information. Different technologies also exert different kinds of influence over the governed. The governor may have *physical influence*, through being in a position to block or use force against the governed actor. The governor might wield *economic influence*, stemming from its ability to allocate material resources, or *social influence*, the ability to alter the governed actor’s relations with other actors. The governor may be able to persuade

¹⁵ Andreas Fischer-Lescano and Gunther Teubner, “Regime-Collisions: The Vain Search for Legal Unity in the Fragmentation of Global Law,” *Michigan Journal of International Law* 25 (2004): 999–1046.

¹⁶ Cf. Scott Camazine et al., *Self-Organization in Biological Systems* (Princeton: Princeton University Press, 2001).

¹⁷ Bruno Latour, “Networks, Societies, Spheres: Reflections of an Actor-Network Theorist” (n. 4).

¹⁸ See, e.g., Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate* (New York: Oxford University Press, 1992); John Braithwaite et al., “Can Regulation and Governance Make a Difference” (n. 4): 1–7; Kenneth W. Abbott and Duncan Snidal, “Strengthening Regulation Through Transnational New Governance: Overcoming the Orchestration Deficit,” *Virginia Journal of International Law* 54 (2009): 501–78.

¹⁹ Peter N. Grabosky, “Using Non-Governmental Resources to Foster Regulatory Compliance,” *Governance* 8 (1995): 527–50.

²⁰ Miller and Rose, *Governing the Present* (n. 3).

²¹ See also Porter, *Trust in Numbers*; Espeland and Stevens, “A Sociology of Quantification” (n. 5).

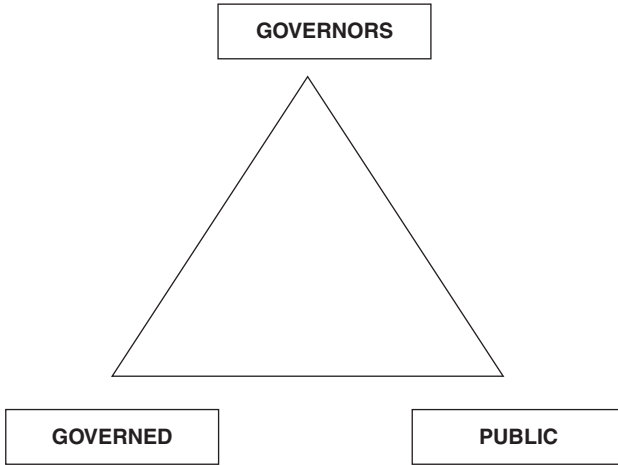


Fig. 1.1 A model of governance

the other actor of the merits of a certain course of action due to being perceived to have special insight which might be termed *scientific expertise* or *moral expertise*. Finally, different technologies of governance may be more or less amenable to particular forms of contestation or subject to different forms of regulation. So, for instance, financial auditing as a technology of corporate governance may be influenced especially strongly by a combination of legal regulation and detailed self-regulation, while environmental auditing is shaped by pressures from a more diffuse set of actors articulating less detailed norms.²²

The term global governance is used in this book simply to denote governance beyond a single state. The governmental agencies of a state are often subject to governance conducted, at least in part, by entities outside the state. These entities may be inter-governmental organizations, hybrid public–private organizations, non-governmental commercial or non-commercial organizations, or other states. The ways in which such governance operates are often immensely intricate, creating substantial empirical and analytical challenges in efforts to understand the roles of indicators as a technology of such governance.

b. Possible effects of indicators on global governance

The use of indicators as a technology of global governance can be expected to affect: where, by whom, and in relation to whom governance takes place (the “topology of governance”); the processes through which standards are set; the processes through which decisions are made about the application of standards to particular cases; and the means and the dynamics of contesting and regulating exercises of power in

²² Michael Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997).

global governance. In the sub-sections that follow we elaborate on each of these claims.

1. Topology of governance

The idea that indicators and other quantitative ways of representing social phenomena can serve as technologies of governance has distinctive implications for the topology of global governance. Indicators are one of the technologies of “government at a distance” (Miller and Rose), allowing certain actors to exercise influence over the conduct of large numbers of geographically dispersed actors, that are readily adapted to forms of governance outside or reaching across distances beyond the state. In particular settings of global governance, using indicators as a technology calls for expansion in ordinary political conceptions of who qualifies as a governor, while at the same time complicating models of governance premised on clear distinctions between governors, governed, and others.

Recognizing indicators as a technology of global governance implies that actors who promulgate indicators ought to be counted among the governors, even if they otherwise would not be recognized as wielders of power in global governance, or would be only to a lesser extent. Thus indicators help constitute or embed power relations. Moreover, simple producers of indicators used in global governance, or actors whose decisions have a significant impact on the form or content of such indicators, may exercise power even where they are not the formal promulgators or users of the indicator.

Including producers of indicators in the class of governors does not mean that tracing the strands of agency and power relations is necessarily straightforward. While in some cases (such as between credit rating agencies and their clients who pay to be rated) there is a symbiotic relationship between those who measure and those who are measured, particularly when the measured entity actively consents to the measuring, in other cases the measurer unilaterally exercises power over the measured. These complex and variegated power relations do not map neatly onto the distinction between governors and governed.

Another complicating factor is that the production of the indicators used in global governance is often a collective process. In many cases promulgators attach their names to indicators whose production involves contributions from a number of other actors. For example, reports and rankings for the Programme of International Student Assessment (PISA) are promulgated by the OECD, but are actually prepared and produced by an Australian consultancy under a contract with the OECD.²³ Moreover, the promulgators of indicators typically rely on data collected by a large network of independent actors stretching from international agencies to national statistical agencies, to local and national NGOs, to villages and local communities. They also rely upon analytical techniques generated by some segment of the scientific community. Consequently, the promulgator of an indicator

²³ Armin von Bogdandy and Matthias Goldman, “Taming and Framing Indicators: A Legal Reconstruction of the OECD’s Programme for International Student Assessment (PISA)”, this volume pp. 52–85.

may or may not be the actor most involved in determining its content. Instead, the promulgator is often more like the “manufacturer” of a consumer product, whose main contribution is to lend its brand name and perhaps its design and marketing expertise and quality control power to the collective product of a global supply chain.

The production of indicators also draws into the practice of global governance, through their own use of the indicators, people who would otherwise be regarded simply as members of the public. For example, when the United States State Department publishes its annual glossy report with indicators of countries’ compliance with anti-trafficking standards, these can be read by activist groups who may influence economic agents such as prospective tourists in Toronto, just as easily as they can be read by government officials. Learning of Costa Rica’s low score may lead a Toronto resident to alter her perceptions of Costa Rica, a country downgraded to the Tier 2 “watch list” in 2011.²⁴ In particular her travel decisions, in combination with the decisions of other members of the public, may (hypothetically) have a material effect on Costa Rica’s tourism revenues.

Indicators may also play significant roles in global governance in helping to constitute actors and shape identities. Some organizations, such as Freedom House or Transparency International, depend for their prominence and influence primarily on indicators they produce. In many organizations, indicator production is important to the business model, helping generate website traffic or a demand for the organization’s consultancy services; some indicators are sold commercially (see the discussion of the economics of indicators below). Disparate actors in different categories may become linked through an indicator which they help construct, or which measures behavior they are concerned with. Indicators may thus play roles in shaping highly decentralized or non-formal governance structures such as networks.²⁵ Indicators may be important in such governance modalities even where no clear delineation of governors and governed and interested public can be made, and where clear overarching human intentionality is lacking or where structural or non-animate elements (such as technological elements) greatly shape outcomes.²⁶

The use of indicators in global governance enhances the role played in global governance by the subset of the public that comprises the scientific community. The scientific community determines the scientific authority of an indicator, which in turn may affect the extent of the indicator’s influence. Producers of indicators are well aware of this fact. For example, Kaufmann and Kraay assert that their World Governance Indicators are more reliable because they are published in scientific journals and peer-reviewed.²⁷ Indicators typically rest on claims to objectivity and social science knowledge, but they differ significantly in the extent to which they reflect social science research and analysis. There are close relations between

²⁴ United States Department of State, *Trafficking in Persons Report*, June 2011. Washington, DC.

²⁵ Kevin E. Davis and Benedict Kingsbury, *Indicators as Interventions: Pitfalls and Prospects in Supporting Development Initiatives* (New York: Rockefeller Foundation, 2011).

²⁶ Latour, “Networks, Societies, Spheres: Reflections of an Actor-Network Theorist” (n. 4).

²⁷ Daniel Kaufmann et al., “Governance Matters”, World Bank Policy Research Paper No. 2196 (1999) 32; see also Daniel Kaufmann et al., “Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996–2008,” World Bank Policy Research Paper No. 4978 (2009).

indicators developed for social science theory testing and those which address policy questions, with the data and analysis of one informing the other.

2. *Standard-setting*

As we have explained, indicators are standard-setting instruments. But while the processes which generate indicators ultimately result in the production of specific goals and targets against which societies are measured, they may be different from other more politically explicit standard-setting processes.²⁸ Whereas political efforts to formulate norms and standards, for example in multilateral inter-governmental negotiations conducted by diplomats, tend to involve processes such as voting or interest-group bargaining or the exercise of material power, the processes in specialist agencies and expert meetings where the standards embedded in indicators are produced, accepted, and supported tend to involve derivation of power from scientific knowledge. As the awareness or the significance of indicators as standards rises, indicator design and production are likely to become increasingly subject to demands made of other standard-setting processes, including demands for transparency, participation, reason-giving, and review.²⁹

Because indicators are by stipulation capable of being used in evaluation, they frequently blend standard-setting with evaluation, by conveying information such as a ranking of the state's performance relative to that of other states and a direction of change in the state's relative or absolute performance by comparison to previous iterations of the indicator. This has the potential to intensify demands for "due process," especially within inter-governmental bodies, as each specific ranked entity has a direct focused interest, going beyond the general interest in good standards, which it may regard as conferring "standing" to raise a challenge.

3. *Decision-making*

In the practice of global governance, many decisions by governing entities are in some way influenced by indicators, although few rely entirely and mechanically on indicators. In the most straightforward case, an indicator promulgated by an extra-national entity is then used by that entity in generating or allocating resources or in influencing behavior. This is, for example, what the World Bank does in promulgating "good governance" indicators that are used by the World Bank itself in deciding how to allocate aid. A modest extension of this occurs where one entity's indicators are used for governance purposes by other entities in the same sector, as when the MCC uses World Bank indicators. A more subtle case arises where the promulgation of the indicator by an extra-national entity spurs demands and

²⁸ The comparison between indicators and standard-setting is explored in Tim Büthe's chapter in this volume. See also Lampland and Star, *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life* (n. 7); Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011).

²⁹ Benedict Kingsbury et al., "The Emergence of Global Administrative Law" (n. 1); see also the chapter by Bogdandy and Goldmann in this volume, pp. 52–85.

governance-related action by diffuse but nonetheless influential groups of other actors. For instance, the World Bank claims that it has prompted many countries to reform their legal systems simply by promulgating and promoting its country-level indicators on the ease of doing business.³⁰ The US State Department's Trafficking in Persons Report claims that it has fostered national anti-trafficking legislation. The regulatory influence of these indicators does not stem exclusively from the ways in which they are used by the World Bank or other development agencies, but also from the ways in which they are expected to be used in lobbying and decision-making by local political constituencies or prospective foreign investors. This shades into a further scenario, in which the indicators have regulatory effects primarily because they have been embraced as guides to appropriate conduct by actors within the state shaping national governmental decisions on national governance. The majority of prominent indicators appear to operate in global governance in even more diffuse ways than this, by influencing professional, public, and political opinion to craft new approaches or take different policy orientations.

Indicators are attractive to decision-makers and designers of decision-making processes because decision-making processes that rely on indicators can be presented as efficient, consistent, transparent, scientific, and impartial.³¹ It is difficult to say which of these factors is most important in any given context. Efficiency and consistency may be factors of special importance in high-volume decision-making; transparency, scientific authority, and impartiality are considerations relevant to the use of indicators in both standard-setting and decision-making, although special issues arise in decision-making.

i. Efficiency

The use of easily-produced or already-available indicators (which simplify more complex and unruly information) is likely to reduce the burden of processing information in the course of decision-making. In principle therefore, reliance on indicators should reduce the time, money, expertise, and other resources required to make decisions. One of the appeals of an indicator technology for human rights treaty bodies is to help in coping with the growing burden of processing country reports as the number of reports increases. On the other hand, selecting or amalgamating among a high volume of different indicators requires expertise and can be costly. It may be viable and attractive for a sophisticated organization. But a multiplication of indicators, some poorly grounded and some extensively marketed,

³⁰ The *Doing Business* team claim that from its first publication in 2003 through to mid-2010, "Governments have reported more than 270 business regulation reforms inspired or informed by *Doing Business*" (*Doing Business Report 2010* (Washington DC: World Bank/IFC 2010), vi). The *Doing Business* team provides assistance to states in how to alter laws or practices to move up the rankings on relevant indicators, and the *Doing Business* team arranges a celebration each year for leaders of "the top 10 reforming governments" based on their improvements documented in that year's report (led in 2009 by Azerbaijan, Albania, Kyrgyz Republic, Belarus, and Senegal). Some 25 countries have inter-ministerial or other committees specifically aimed at "improving the business environment" using the *Doing Business* indicators as one guide (*ibid.*, 15). The World Bank Group's newer *Investing Across Borders* team pursues similar approaches designed to increase the influence of its indicators.

³¹ Porter refers to these virtues compendiously as "objectivity." Porter, *Trust in Numbers* (n. 5).

may lead to confusion and worse decision-making for other organizations and their constituencies.

The cost–benefit attractions of relying on indicators are particularly pronounced when sophisticated numerical data and information processing technology are readily available. It seems likely that the expansion in the use of global indicators since the 1990s is linked to the increasing accessibility and quality of social and economic statistics, the ever-declining cost of computing, as well as improvements in and dissemination of statistical techniques. National statistical systems are generally in a long-run pattern of improvement (although there are also cases of decline in governmental statistical systems). For example, developers of an indicator for the right to health were already able, by the early 2000s, to present data on 72 indicators for 194 countries using data available on the Internet.³²

In some contexts, the quality of indicators may actually be a function of the total supply of indicators because some indicators are arguably most useful when aggregated with other similar indicators.³³ This raises the intriguing possibility that, at least for relatively sophisticated actors, the use of indicators may be a self-reinforcing phenomenon: as more indicators are produced, aggregations of indicators become more reliable, more indicators are used, more indicators are produced, et cetera. Greater supply of indicators also influences the ecology of indicators, with comparisons among them enabling selection of the most robust and reliable, and possibilities of continuous improvement.

It seems plausible that reducing the costs of decision-making becomes more attractive (sometimes even imperative) as the amount of decision-making and the need for rapid decisions increases. Thus, the striking increase over the decades since 1990 in the creation and use of indicators as forms of knowledge for global governance arguably reflects the greater demand for readily available and easily used comparative knowledge to inform decision-making as well as the increasing supply of information. The reliance on indicators in global governance seems to be associated with developments such as increases in population and in levels of economic activity, which in turn determine the scale and intensity of social and economic interactions susceptible to governance, and with specific institutional developments affecting the nature of governance decision-making.

ii. Consistency

To the extent that indicators provide unequivocal ordinal data, they can be translated into numerical form and used as inputs into decisions made in accordance with rules expressed in mathematical form (such as “approve the grant if $(A * B)/C > 3$ ”). A distinctive feature of rules that can be expressed as these kinds of mathematical operations is that they yield consistent results; given the same inputs, the output will be the same regardless of who is applying the rule or when it is being applied. Holding this process constant also enables consistency over time.

³² Gunilla Backman et al., “Health Systems and the Right to Health: An Assessment of 194 countries,” *The Lancet* 372 (2008): 2047–85.

³³ Kaufmann et al., “Governance Matters” (n. 27).

Consistency is likely to increase the legal or moral authority of decision-making in some contexts.

iii. Transparency

The simplicity of indicators makes it relatively easy to communicate them to third parties. This is significant whenever an effort is made to give third parties access to the informational basis for a decision; it should be relatively easy to communicate the basis for a decision based on indicators. This transparency can be superficial because the raw data used to construct indicators, and the methods used to simplify those data, are not necessarily easy to communicate and may in fact be treated as confidential. Even when such detailed information is provided, users may well not delve into the complexities and limitations of the underlying data and the analytic choices made in converting it into an indicator.

iv. Scientific authority

Reliance on indicators has the potential to displace unmediated subjective data and replace it with data whose relevance and reliability has been endorsed, to some extent, by a community of scientists. This in turn means that the credibility of decisions based on indicators can depend in part on the extent to which the indicator is seen to be endorsed by various scientific communities and the amount of authority commanded by those communities. An indicator may gain credibility from its association with particularly prominent individual scholars.

v. Impartiality

Basing governance decisions solely on publicly disseminated indicators excludes the possibility of basing them on subjective considerations of, or private data known only to, the particular decision-maker. As Porter has argued, the less a governor is trusted the more appealing this kind of demonstrated impartiality becomes.³⁴ This impartiality is limited, however. The reasons for simplifying raw data in one way instead of another, or choosing to rely upon one indicator rather than another, may be highly subjective. The decision-maker may be involved in this process, whether by constructing the indicator, determining itself which indicator to use, or signaling a demand for an indicator conforming to its preferences which a supplier then meets. It is in any case almost inevitable that indicators are shaped by the knowledge and experience of the experts who produce them. This knowledge and experience may in some cases be dominated by that of the first movers or early adopters of quantification in a particular area of social policy.

4. Contestation

A great deal remains to be learned about when, how and why the governed (or rival governors) contest the use of indicators, but we expect it to take both general and

³⁴ Porter, *Trust in Numbers* (n. 5).

long-established forms such as lobbying and litigation, as well as distinctive forms that are especially suited to changing or resisting governance through indicators, such as refusal to participate in data collection, challenges to scientific validity, or creation of alternative indicators. Contestation can take the form of debates about the data used and not used in indicators, the weightings criteria for the indicators, or about the embedded social and political theory of the indicator. Contestation strategies can include the creation of new indicators, and resistance to or discrediting of existing indicators and their producers or users. This may in turn prompt modifications to the indicator, or counter-strategies by producers and users.

Because they obscure the socio-political theoretical claims embedded in their construction, the use of indicators can make it relatively difficult to contest the use of those theories in global governance. Indicators may mask large areas of missing or incomplete data, inability to draw significant distinctions between entities that are nonetheless hierarchically ranked, much higher levels of underlying uncertainty than the indicator depicts, and choices about weighting of different components of composite indicators (which in some notable privately produced indicators are not disclosed at all). On the other hand, those with special expertise in the construction or analysis of indicators can overcome these impediments to technical contestation and exercise greater influence than they could in purely political settings. Limitations in the ability to contest the exercise of power by global decision-makers tend to shift the balance of power toward “technical” experts, that is to say, people with expertise in the construction or analysis of indicators.

The rapid growth in prominent indicators in global governance is a time-compressed phenomenon that was initially sufficiently circumscribed for case studies about early patterns of debate, acceptance and challenge to be used as a basis for some cautious generalizations. Several of the most prominent indicators in global governance began as efforts led and shaped by social-science communities. Eventual “scientific” acceptance of these indicators can be traced back to a time before “public” knowledge is settled about the issue through various controversies and challenges. The development of these indicators instantiates, to some extent, the process of developing scientific knowledge described by Latour.³⁵ Like other forms of scientific knowledge production, an indicator builds on existing concepts, techniques, and categories of understanding that are taken for granted as correct, and on networks of experts. Indicators are shaped both by technical factors, such as the statistical properties of an indicator when compared to other indicators, and social factors, such as social networks, perceived expertise, relational interactions, institutions, and allies.³⁶ These processes are collective and take place over time. Once the indicator has been established with wide scientific support (even while continuous scientific debate and refinement remains part of the further life of the indicator), a process of wider public acceptance occurs, as networks of actors and institutions adopt the indicator and consequently increase its credibility and

³⁵ Latour, *Science in Action: How to Follow Scientists and Engineers Through Society* (n. 7).

³⁶ *Ibid.*, 29.

legitimacy, perhaps even converting it into a standard against which other indicators are evaluated.

The results of a survey we conducted of reporting about four major indicators—UNDP's HDI, Transparency International's Corruption Perception Index,³⁷ Freedom House's Freedom in the World indicator,³⁸ and the World Bank's Doing Business indicators—in three major US and UK opinion-shaping newspapers and magazines are consistent with this model.³⁹ In the first year or two after an indicator is released, there is discussion and debate about the indicator itself, but after a few years, the indicator is presented in these news media largely as a fact that describes a country's situation, with virtually no discussion about the source of the data or the nature of the indicator itself. In some cases, however, such as opposition from organized labour groups and the International Labour Organization to the 'Employing Workers' component of the Doing Business indicators, efforts at contestation can be intense. As such indicators have become more and more significant as technologies of global governance, the stakes of contestation have risen. Producers of indicators who may have viewed themselves as scientists or technicians working outside political and legal arenas, have been drawn into highly political conflicts.

5. Regulation

One outcome of contestation can be demands for regulation of indicator-related processes and activities. Some of these demands instantiate general patterns of demands for increased transparency, reason-giving, participation, review, and accountability in global governance institutions and processes, particularly where public authority is being exercised, but also in relation to some private governance actors.⁴⁰ Given the distinctive features of indicators as a technology of global governance, we expect growth in specifically adapted proposals or efforts to regulate indicators. These may take a variety of forms, several of which have been manifested in debates about regulation relating to sovereign debt ratings by credit rating agencies.⁴¹ For instance, *producers of indicators* could be subject to scrutiny (although not necessarily legal obligations) with reference to human rights standards, domestic constitutional norms, and principles of global administrative law. Others may be regulated in the same ways as private actors such as multinational corpora-

³⁷ Transparency International, Corruption Perceptions Index 2010, available at <http://www.transparency.org/policy_research/surveys_indices/cpi/2010/results> (accessed September 30, 2011).

³⁸ Freedom House, *Freedom in the World 2010* (Washington, DC: Freedom House, 2010).

³⁹ This survey examined news stories in the *New York Times*, the *Washington Post*, and the *Economist* and compared coverage in the first year after each indicator was created with coverage in 2004 and 2009. We thank Jessica Shimmin for work on this.

⁴⁰ Kingsbury et al., "The Emergence of Global Administrative Law" (n. 1); Benedict Kingsbury, "The Concept of 'Law' in Global Administrative Law" (n. 1); Armin von Bogdandy et al. (eds), *The Exercise of Public Authority by International Institutions: Advancing International Institutional Law* (Heidelberg: Springer, 2009).

⁴¹ Amadou N.R. Sy, "The Systemic Regulation of Credit Rating Agencies and Rated Markets," IMF Working Paper WP/09/129 (Washington, DC: International Monetary Fund, 2009); Mathias Audit, "Aspects internationaux de la responsabilité des agences de notation," *Revue critique de droit international privé* 100 (2011): 581–602.

tions or networks of firms linked by transnational supply chains. These analogies suggest also the possible relevance of regulatory mechanisms such as competition law, transnational tort claims, and self-regulation. Procedural obligations on producers might require them to be transparent about the methods used to produce indicators and their limitations, to allow interested parties to participate in some way in the design process, and/or to accept some accountability in problematic cases for effects on external actors. Alternatively, producers might find their indicators held to externally administered standards of reliability and validity. Finally, structural interventions might be designed to foster healthy competition among producers. So, for example, public bodies might support or subsidize the production of competing indicators, or certain organizations already exercising other substantial powers as governors might be encouraged to refrain from promulgating indicators.

Other regulatory interventions might target the *users of indicators*. For example, use of indicators in global governance may spawn systematic efforts to educate users of indicators, and the members of the public who confer authority upon them, about both the costs and benefits associated with using indicators. Alternatively, regulation could focus on empowering *actors who are governed by indicators* by, for example, giving them access to the scientific expertise they need to contest decisions based upon indicators.

IV. Outline of this volume

Using case studies of indicators that are influential in global governance—university rankings, the OECD’s PISA rankings of achievement among cohorts of school students, the UN Global Compact and other indicators of “corporate social responsibility,” measures of state fragility and state failure used differently by aid agencies and the military, the World Bank’s influential neo-liberal “Doing Business” and “Investing Across Borders” global rankings, ratings of companies for social impact investors, indicators used to measure police performance, indicators measuring health systems, trafficking in persons, internal displacement in Colombia, and the performance of humanitarian agencies in Haiti—the chapters illuminate some of the hypotheses proposed in the previous section about the ways in which indicators operate and affect global governance. In aggregate, they focus particular attention on the interplay of power and authority, dynamics of global and local indicators, and relations between indicators and law, including human rights law and global administrative law.

The chapters comprising the remainder of Part I of the book consider different theoretical approaches to the study of indicators.

Tim Büthe constructs a political-economy model of the supply and demand for indicators, recognizing, however, that most indicators are supplied and used in political contexts which do not closely follow market precepts, so that supply and demand may well be in long-term disequilibrium. He emphasizes the insufficiency of focusing simply on the producers (suppliers) and users (demanders) of indicators,

and accordingly formulates a model that identifies four relevant sets of stakeholders. In addition to those who would conventionally be called producers and users his model encompasses (a) actors who call for or provide material support for the production of indicators, regardless of whether they actually produce or use the indicators; and (b) the targets of the rules, “the subset of the users who either are themselves being measured or who act and speak on behalf of the objects or institutions being measured.” Drawing on several other chapters in the book, Bütte applies his model particularly to private producers of indicators, such as the law school rankings produced by the for-profit entity US News and World Report, and the Freedom in the World index produced by the NGO Freedom House.

A public law approach to global governance indicators is taken by Armin von Bogdandy and Matthias Goldmann. Drawing on German public law thinking as well as theories of public law beyond the state, they propose that the promulgation of the PISA evaluations and rankings by the OECD be characterized in legal terms as a “standard instrument,” in this case a “national policy assessment” (NPA). Because the OECD can be regarded as exercising international public authority, its NPAs would be subject to public law requirements, such as that the content and political direction of the NPA program be clearly defined under a mandate with a definite legal basis, adopted through a multi-stakeholder process, and implemented by a geographically balanced group of experts selected under an open process. The producers of NPAs would be required to give reasons for specific actions, to respond to criticism, and to maintain independence from national governments by which they might be captured.

The sociological approach taken by Wendy Espeland and Michael Sauder examines the processes by which a particular indicator comes to have meaning, effects, and uses in ways and places very different from those envisaged in its original production. This “dynamism” of an indicator depends on who notices or ignores the indicator, how different constituents use it, and how these users are connected to one another. “Successful” indicators become significant or authoritative as they accumulate networks of constituents, technologies, and things. The case studies in this chapter are of three sets of indicators of higher educational institutions: US law school rankings, transnational rankings of business schools, and world university rankings. Espeland and Sauder, like Bütte, highlight particular social and psychological factors that give special power to numbers and rankings as definite, transitive, simplifying heuristics.

Actor-network theory (material semiotics) inspires the exploration by Ronen Shamir and Dana Weiss of the emergence of measurable and comparable representations of “corporate human-rights responsibility.” The authors consider symbolic representation of indicators in maps, diagrams, and social-branding labels, arguing that indicators tend to generate secondary and even third-tiered indicators (“indicators of indicators”). The second layer of analysis explores “corporate human rights responsibility” as a social performance which is simultaneously enacted through two interacting social modalities: regions and networks. The authors emphasize the extent to which indicators are actors whose work simultaneously

assembles the net of “corporate human rights responsibility” and the compilations that produce regions. Indicators are double agents: working to preserve the integrity of the network but also enabling the production of regions. In addition, the authors identify and criticize the practice of indicators moving “corporate human rights responsibility” further and further away from addressing or ameliorating the real risk-to-people.

A genealogical approach, producing a kind of “history of the present” is taken by Nehal Bhuta in his study of the USAID state fragility index as an example of the construction, uses, and purported measurement through indicators of the concepts of fragile, failing, or failed states. This terminology, he comments, operates as “a ‘trading language’ used to talk about complex social realities which are highly heterogeneous and about which there is little theoretical and empirical agreement. Once the concept is pragmatically and provisionally in use—and used in a variety of diagnostic, prescriptive and evaluative ways—its consistency with any ‘outer reality’ is less significant for its users than its correspondence with a (porous) set of acceptable uses and deployments.”

Part II of the book uses social science methodologies to investigate how power and authority are exercised by and through indicators.

Katharina Pistor traces the history of the first generation of the indicators of governmental institutional quality—global indicators bearing labels such as “rule of law,” “security of property rights,” and “bureaucratic efficiency.” Data collected and disseminated by private for-profit groups or by Freedom House was used in the mid-1990s and transposed by economists into measures correlated with development success under the Washington Consensus mantra that “institutions matter.” Pistor argues that it is not the creation of the indicators that is problematic but rather the transposition of indicators that were designed for a narrow purpose into justifications for large-scale development policies by leading multilateral agencies, where neither the assumptions of policymakers that led them to these data nor the re-interpretation of the data is further scrutinized. She urges that raw data be made readily available, and alternative data sets used, to challenge existing assumptions rather than simply seeking to validate them and the policy choices with which they are associated.

Terence Halliday presents the results of detailed empirical studies into efforts by inter-governmental financial institutions (IFIs) to measure the quality of national laws on matters such as commercial law and corporate bankruptcy, and the use of these legal assessments in promoting change through inter-country comparison (especially countries in the same region) and appraising change through inter-temporal comparisons of a single country. He situates these initiatives within an overall ecology of partly competitive interactions between IFIs, but also within IFIs. The latter related particularly to the roles and aspirations of IFI legal departments as they moved from simply providing legal services to engaging in development policy under the new theory that “good” (reformed) national legal institutions were important to development and to financial system resilience. Halliday argues that the ecological, organizational, and professional challenges faced by IFIs and their legal staffs set significant limits to changing the uses of indicators and shifting

to non-indicator or less prescriptive methodologies, however desirable such changes or shifts might be.

Angelina Fisher studies the ways in which the WHO/UNICEF immunization coverage indicators have been used, at times extending far beyond their purported scope. The Global Alliance for Vaccines and Immunization (GAVI), for example, uses the WHO/UNESCO DTP3 vaccine coverage indicator to determine the type of funding it will make available to a country. Under GAVI's rules, countries with less than 70 percent DTP3 coverage of children at age one year are *not* eligible for funding to support introduction of new vaccines. In the Heavily Indebted Poor Countries debt relief initiative, this indicator is used as a requirement for the country to show progress on good governance. In other situations, these indicators are used in effect as proxy measures for the overall quality of national health care systems. As Fisher explains, this situation may be explained in part by the absence of any other single indicator of national health system performance suitable for making cross-country comparisons. In 2000, WHO published *The World Health Report 2000—Health Systems: Improving Performance*, measuring health systems in 191 member states on the basis of five indicators: overall level of population health; health inequalities (or disparities) within the population; overall level of health system responsiveness (a combination of patient satisfaction and how well the system acts); distribution of responsiveness within the population (how well people of varying economic status find that they are served by the health system); and the distribution of the health system's financial burden within the population (who pays the costs). The political uproar that followed (prompted not the least by the US ranking 37th) caused WHO to stop any subsequent explicit ranking of health systems. Immunization coverage indicators have filled the gap.

Part III of the book explores the different dynamics of global, regional, and local indicators, focusing on ways in which these acquire authority and significance, and influence policy process.⁴²

Studying the roles of indicators of the numbers and circumstances of internally displaced persons in Colombia, an issue of tremendous importance in Colombia as a consequence of massive violence and attempts to remediate its effects, René Urueña traces the movement of numbers from local compilers with little direct influence, into reports of prestigious external bodies such as the United Nations, then back into the Colombian news media and political debates. He argues that there is a dialectic element to the process of creating and applying the indicators. Indicators may be creatively appropriated for local purposes that are not necessarily connected with their original "global" origin, and both governments and non-state actors adopt indicators as part of their rhetoric. Urueña shows that those who are measured will try to influence the measurement, making indicators an important aspect of their strategic political choices. The complex interaction between the producers and users of indicators cautions against considering governance through indicators as a one-way, top-to-bottom process.

⁴² This is also a major theme in the international collaborative research project directed by Merry, Kingsbury, and Davis and funded by the National Science Foundation.

Christopher Stone makes a normative argument for bottom-up generation of locally usable near real-time “active indicators,” as preferable to top-down production of global indicators or donor-demanded indicators which engross external constituencies but have little real local salience in direct management. Studying indicators of different performance patterns across police divisions in Jamaica, or different drivers of pre-trial detention in a Nigerian prison, he argues for iterative development of indicators from these local experiences into inter-country dialogues among the relevant professionals, and thence perhaps the eventual aggregation of these into global indicators.

Chapters in Part IV of the book consider the relations between law, human rights values, and indicators in human rights and humanitarian governance.

AnnJanette Rosga and Margaret Satterthwaite point out that efforts to use indicators within the *law* of international human rights are inevitably criticized for the kinds of pathologies identified in the literature on national audit practices: quantifying the not-adequately-quantifiable; missing data and concoction of data; disguising the processes of politics and judgment; the measures becoming endogenous as targets; the frequent need to revise what is measured, undercutting inter-temporal comparability. Nonetheless, they see a growing potential for suitably tempered indicators to play valuable roles. A project of the UN Office of the High Commissioner for Human Rights (led by several outside advocate-experts working with the secretariat) from 2005 aimed to produce internationally prescribed indicators for the ICESCR and several major UN human rights treaties. Development of such indicators might help address concerns about the perceived legitimacy of the supervisory committees under these treaties, by shoring up their authority through relying on social science and statistics, and thus making it appear that their assessments of each state were based on quantified analysis applied to all states rather than being specific acts of (political) judgment. Some core difficulties in this turn to indicators have been manifest in this project: problems of authoritatively specifying adequate indicators where the rights themselves are under-specified; the improbability of “structural” and “process” indicators measuring major causal effects on “outcomes” in many cases; and the inability of indicator-based approaches to free the committees from acts of judgment in real cases. The initial aspirations of the OHCHR project to produce universally applicable indicators for assessing compliance and fulfillment of rights and enabling inter-country comparison were subsequently watered down, to a more modest and realistic aim of producing illustrative indicators. Rosga and Satterthwaite welcome this, and suggest that indicators might play a role in helping peoples and publics to exert the kinds of pressures and constraints on governments that human rights advocates have long sought through the rights themselves.

Wariness of governments about this possibility has been one factor slowing the development by inter-governmental bodies of compliance-focused indicators under international treaties against human trafficking. Anne Gallagher and Janie Chuang note this in their chapter, and focus on the major effort to produce such indicators, which has been undertaken by the US State Department pursuant to US legislation. This power is asserted unilaterally; almost no cases exist where an

international institution has delegated this power to a single government, or where other countries ranked have requested such rankings. Such unilateralism may nonetheless receive some support. It can overcome a collective action problem, where all or most governments agree that an indicator is desirable, but they cannot agree on effective criteria, and the decision rules or bargaining arrangements drive them down to the lowest common denominator of no indicator or an insipid one. Unilateralism can also function, albeit crudely, as a form of virtual representation, where the individuals and groups who are intended to benefit from performance of the treaty lack the influence themselves to incentivize governments to perform the treaty or to monitor performance. The US State Department's trafficking in persons indicators are, however, produced under criteria set by US legislation. These criteria are not identical to the relevant international treaty, although broadly similar. Politics within the US have influenced aspects of this monitoring: for example, a strong focus on prostitution under the G.W. Bush administration was attenuated under the Obama administration. Micro-political US concerns may also have influenced some reporting or non-reporting of particular trafficking incidents in the State Department's publications. Overall, however, Gallagher and Chuang suggest that the US unilateral approach to trafficking has probably provided an otherwise under-supplied public good.

Much more disquiet about one effect of global anti-trafficking law, policy, and indicators is expressed by Marina Zaloznaya and John Hagan in their study of the uses of the anti-trafficking agenda by the authoritarian government of Belarus. The authors point to the uses by the Belarus government of anti-trafficking as a basis for making travel abroad by students and young people much more difficult, and for other autarkic and nationalist agendas. They express concern that the State Department trafficking indicators and accompanying narrative take no account of other forms of human rights repression or excesses of policing systems in Belarus, and instead provide legitimating endorsement to the regime's efforts, an endorsement that single-issue international anti-trafficking organizations have also tended to concur in. This study illuminates the important but under-researched theme of the significance of global indicators in authoritarian contexts.

Indicators operate as core elements of systems of regulation in many contexts. Margaret Satterthwaite's chapter on the operation and effects of indicators used to guide and assess the work of (mainly non-governmental) humanitarian relief agencies in their on-the-ground operations in Haiti provides an illuminating example. Two major sets of standards were adopted as a form of self-regulation by groups of major international NGOs after the disastrous problems they faced in Great Lakes camps and in Yugoslav "safe areas" in 1994–95. These are the widely used Sphere indicators, and the Humanitarian Accountability Partnership's standard (although relatively few such NGOs have HAP certification thus far). The Sphere indicators, in particular, are used by major donors to guide decisions on which NGOs to fund, by peer NGOs as standards to hold each other to in order to maintain legitimacy of the whole "industry," and by staff within these NGOs to guide and critique their own performance internally. The processes for setting the standards included in Sphere often resemble regulatory processes. For example, the

standard of 15 liters of water per person per day was set taking into account what was plausibly attainable in many countries, but it was set well above bare survival level in order to provide leverage to relief agencies when seeking water supply funding and when working with host governments so that long-term reconstruction infrastructure water target levels would not be too low. The Sphere indicators have at times had some of the problems resulting from pathologies in incentives that characterize both indicators and regulation: after the January 2010 earthquake in Haiti formal relief camps were not established in areas of Port-au-Prince where it would be impossible to meet the Sphere standards of shelter, water, security, etc.; new Sphere-compliant camps were established in areas of Port-au-Prince where beneficiaries did not want to be; or (in a Sudan case) the number of people in food programs was limited to ensure each participant received the nutrition levels set in an earlier version of the Sphere indicators. The Sphere *Handbook* also has some regulatory silences: it said little (before its 2010 version) on protection of local people (action to prevent arbitrary forced displacement, or ensure their free movement rather than de facto detention), and continued after 2010 to lack indicators to measure performance on these issues, in contrast to the detailed indicators on more straightforward food and shelter functions of humanitarian agencies.

Sarah Dadush presents a study of the development and use of indicators and related reporting systems in social impact investing, through the Global Impact Investment Rating System (GIIRS) and the Impact Reporting and Investment Standards (IRIS). Social issues are coming to be weighed in commercial investments, while market values are increasingly inserted into philanthropy. In this dual process, quantification and the emulation of commercial ratings and accounting make specialized “investment impact” indicators increasingly significant as bridging and blurring devices. Dadush points to positive attributes of this system, but also to some of its costs, including a shift away from local self-expression of distinctive community initiatives in specific developing country settings, toward a standardized structure and language of global markets with which small-scale entities have little choice but to conform.

Part V of the book considers whether (and if so, when and in what form) indicators and indicator-related processes and activities should be subject to regulation.

Nikhil Dutta’s chapter on market mechanisms of regulation focuses on the extent to which indicator producers provide transparency, reasons, participation opportunities, or review mechanisms in their activities. He postulates two means by which levels of these types of accountability in indicator generation are determined where some kind of market for the indicators exists. The first is the Demand Hypothesis, which posits that the response of generators to the demand from users and targets for accuracy determines observed levels of accountability. The second is the Supply Hypothesis, which predicts that indicator generators provide higher levels of accountability in order to attract users. The author tests how well each of these hypotheses explains observed levels of accountability in the generation of three indicators: the International Country Risk Guide ratings produced by the

PRS Group, a for-profit private company; the Freedom in the World ratings generated by Freedom House, a non-profit non-governmental organization; and the Minorities at Risk variables developed by the MAR Project, an academic project at the University of Maryland. The case studies confirm elements of the Supply Hypothesis, but with inflections.

In the final chapter, Sabino Cassese and Lorenzo Casini address arguments for and against the regulation of different global indicators, using the ratings produced by credit rating agencies as a particular illustration. They argue that, in normative terms, the determination as to whether a regulatory framework is needed depends on factors such as the type of indicator at issue, characteristics of the indicator-producing entity, and the kinds of users involved. The chapter sets out a taxonomy of different types, uses, and conditions of use of indicators, in order to distinguish cases in which indicators as accountability-enhancers require protection from regulation, from situations in which some public regulation may be required.

V. Conclusion

A premise of this book is that indicators are a technology of global governance, with distinct properties which we have sought to delineate and specify. Contributors to this volume have engaged in the systematic study of quantification and indicators as a technology of global governance, suggesting schematically the kinds of effects indicators could have on global governance, including on the topology of global governance (who are governors and governed, and in what ways), effects on processes of standard-setting and decision-making, and effects on ways in which contestation of governance occurs. Potential effects on the demand for and the supply of regulation in particular modalities, as well as effects on power and identities are also considered. The chapters in this book illuminate reasons for the growing use of indicators in global governance, the actual effects of particular indicators, and interactions between indicators and other technologies of governance, including law as well as different methods of governance by information.

2

Beyond Supply and Demand: A *Political*-Economic Conceptual Model

Tim Büthe*

Introduction

An indicator, as defined by Davis, Kingsbury, and Merry, identifies and thus circumscribes a particular phenomenon as distinct and separable from its broader social context, then simplifies the often still multi-dimensional phenomenon in purportedly systematic ways so as to allow for numerical (at least ordinal) representation on a single dimension.¹ By turning complex social phenomena such as democracy, freedom, human development, the quality of education, or the stability of states into something on which very different organizations, groups, or entire countries can be—seemingly straightforwardly—compared and ranked, indicators appeal to our desire for clarity and simplicity.² Those who supply indicators achieve this clarity and simplicity—which makes the indicators both powerful and useful—by turning the issue at hand into a measurement problem and then developing a set of explicit, often technical rules for measurement. Measurement standards are therefore at the core of indicators.³

* For comments on previous drafts, I am grateful to Eyal Benvenisti, Cindy Cheng, Kevin E. Davis, Nikhil K. Dutta, Katharina Pistor, Alison Post, Michael Sauder, Richard Stewart, and the participants of the NYU School of Law conference on Indicators as a Technology of Governance in September 2010.

¹ The very name of the indicator may entail both primary and secondary claims about what is measured and what should be measured, as well as who made the original/key intellectual contribution; see also Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, "Indicators as a Technology of Global Governance," *Law and Society Review* 36:1 (2012): 71–104.

² Irving Biederman, H. John Hilton, and John E. Hummel, "Pattern Goodness and Pattern Recognition," in Gregory R. Lockhead and James R. Pomerantz (eds), *The Perception of Structure* (Washington DC: American Psychological Association, 1991), 73–95; Nick Chater, "Simplicity and the Mind: Spearman Medal Award Lecture 1996," *The Psychologist* 10:11 (November 1997): 495–8; Tania Lombrozo, "Simplicity and Probability in Causal Explanation," *Cognitive Psychology* 55:3 (November 2007): 232–57. See also Maxwell J. Charlesworth, "Aristotle's Razor," *Philosophical Studies* 6 (1956): 105–12 and Herbert Marcuse, *One Dimensional Man* (Boston: Beacon Press, 1964).

³ More precisely, any indicator entails a measurement standard or measurement rule; it *may* entirely consist of such a standard. For purposes of this chapter, I will use indicator, measurement standard, and rule as synonyms.

In this chapter, I emphasize the easily obscured relationships of power that are inherent in setting such standards, particularly if the resulting indicators are widely used. I begin with a critique of theoretical approaches that model standards and standard-setting as equilibria of supply and demand, and I suggest an alternative conceptual model that distinguishes four major sets of stakeholders.⁴ I argue that this conceptualization is both normatively and analytically useful for the study of indicators as instruments of global governance. In the second section, I sketch possible explanations for each of the four sets of stakeholders, that is, explanations for the demand for indicators, for the supply of such technical rules, for their implementation or use, and for behavioral adaptations by the measured—whom I call the ultimate “targets” of indicators, even if change in their behavior is not always originally intended. In the final section, I apply the abstract conceptual stakeholder model to two specific indicators prominently discussed in several chapters in this volume: The *US News & World Report*’s law school ranking, and the Freedom in the World ratings, published annually by the Washington-based NGO Freedom House.

Beyond supply and demand: A *political-economic* model of indicators

Providing measurement standards is no minor contribution. Basic measurement standards such as weights and units of volume, for example, have since ancient times been considered a prerequisite for the efficient operation and perhaps even for the very existence of markets.⁵ Agreement on what is today known as the international system of units—the metric standards for measuring the physical world, which define units of measurement for phenomena such as mass, length, time, electric current, and their logical relations to each other—enabled major scientific breakthroughs and technological development for decades.⁶ Measurement stan-

⁴ I first developed the core ideas underpinning this conceptual model in my introduction to a recent special issue of *Business and Politics*. See Tim Büthe, “Private Regulation in the Global Economy: A (P) Review,” *Business and Politics* 12:3 (October 2010). Those ideas are here adapted specifically to the analysis of the kinds of measurement standards that are at the core of indicators.

⁵ See, e.g., Hendrik Spruyt, “The Supply and Demand of Governance in Standard-Setting: Insights from the Past,” *Journal of European Public Policy* 8:3 (June 2001): 371–91 and Wolfgang Trapp and Heinz Wallerus, *Handbuch der Maße, Zahlen, Gewichte und der Zeitrechnung* (5th edn, Stuttgart: Philipp Reclam Junior, 2006). For examples of the role of such standards in the international integration of markets, see Tim Büthe and Walter Mattli, *New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011), esp. chs 4 and 6.

⁶ For a discussion of the role of the international system of units in the development of electro-technology, see Tim Büthe, “The Power of Norms; the Norms of Power: Who Governs International Electrical and Electronic Technology?” in Deborah D. Avant, Martha Finnemore, and Susan K. Sell (eds), *Who Governs the Globe?* (New York: Cambridge University Press, 2010), 292–332, esp. 324ff; and Hans Teichmann, *Celebrating the Centenary of SI—Giovanni Giorgi’s Contribution and the Role of the IEC* (Geneva: IEC, 2001). For an overview of the broader literature on the role of such standards in science and technology, see Andrew L. Russell, “Standardization in History: A Review Essay with an Eye to the Future,” in Sherrie Bolin (ed.), *The Standards Edge* (Ann Arbor, MI: Sheridan Press, 2005), 247–60.

dards thus often bring benefits that are not diminished (and may even be increased) by widespread use, a characteristic to which economists refer as being “non-rival in consumption.” In addition, users of measurement standards often cannot be “excluded,” i.e., users cannot easily be made to contribute to the costs of supplying a measurement standard—costs that can be substantial if developing the standard requires specialized expertise.⁷ In sum, measurement standards have the two key characteristics of public goods, which tend to be undersupplied.⁸ The supply of indicators therefore has to be explained, all the more when they are supplied—like many of the indicators covered in this volume—voluntarily by private bodies.

At the same time, the “demand” for any particular measurement standard should not be taken for granted, despite the public good characteristics of such standards. Setting a particular metric as *the* standard entails—and, depending on one’s perspective, brings closure to or forecloses—conflicts of interest, even if the standard may appear innocuous, aimed at measuring inanimate objects based on fundamental scientific principles. In the case of the international system of units mentioned above, Volt and Ampere, for instance, were chosen from among numerous alternative units of potential electromotive force and current, respectively, often with substantial commercial stakes attached to them.⁹

In a review of the early literature on private governance, I therefore argued that the global governance literature needed to explain both the supply of, and the demand for, inter- or transnational rules.¹⁰ In fact, scholars of international political economy have provided several such models of supply and demand. Mattli and Woods, for example, use a model of supply and demand to assess the normative desirability of regulation based on the likelihood that such regulation serves the public interest (which itself is a meta-level measurement standard).¹¹

Such models of supply and demand, however, have two major shortcomings. First, they implicitly assume that three subsets of stakeholders of any particular indicator perfectly overlap, which allows the analyst to omit the distinction among them and to subsume them among those who “demand” these measurement standards for the particular phenomenon at hand. In effect, this renders some of

⁷ Beyond initially developing such technical rules, maintaining/updating and institutionalizing them is costly; see, e.g., Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton: Princeton University Press, 1984).

⁸ For a discussion of this issue with specific application to standard-setting, see Charles P. Kindleberger, “Standards as Public, Collective and Private Goods,” *Kyklos* 36:3 (September 1983): 377–96.

⁹ Tim Büthe, “Engineering Uncontestedness? The Origins and Institutional Development of the International Electrotechnical Commission (IEC),” *Business and Politics* 12:3 (October 2010) and Büthe, “The Power of Norms; the Norms of Power” (n. 6).

¹⁰ Tim Büthe, “Governance through Private Authority? Non-State Actors in World Politics,” *Journal of International Affairs* 58:1 (Fall 2004): 281–90.

¹¹ Walter Mattli and Ngaire Woods, “In Whose Benefit? Explaining Regulatory Change in Global Politics,” in Mattli and Woods (eds), *The Politics of Global Regulation* (Princeton: Princeton University Press, 2009), 1–43. For other examples of such models of supply and demand, see, e.g., Donald J. Lecraw, “Some Economic Effects of Standards,” *Applied Economics* 16:4 (August 1984): 507–22 and Shyam Sunder, “Regulatory Competition Among Accounting Standards Within and Across International Boundaries,” *Journal of Accounting and Public Policy* 21:3 (Autumn 2002): 219–34.

the stakeholders and the inherently *political nature* of indicators invisible. Second, they suggest an equilibrium without necessarily addressing how it might be reached, even when the realm of global governance hardly operates like a market. In effect, this renders the *political process* of establishing global rules invisible. I therefore argue for moving beyond models of supply and demand to a more differentiated conceptualization that recognizes the three-plus-one analytically distinct groups depicted in Figure 2.1. I submit that this broader conceptualization of the stakeholders invites theoretical and empirical analysis of the extent to which the groups overlap and thus makes visible the relationships of power that exist whenever indicators play a role in global governance.

Specifically, I argue for distinguishing four core groups of stakeholders. The first group consists of socio-political actors who either overtly call for indicators or value them to the point where they are willing to give credit or pay some cost for their provision. The intensity of their preferences explains what may be called the level of demand for such measurement standards for any particular issue. I call them the “rule-demanders.”¹² The second group consists of the actors who write, maintain, and disseminate such measurement standards. These are the “rule-makers.” As noted above, their supply needs to be explained because these activities are costly. Actual or potential “users” constitute the third group. This group consists, specifically, of those who utilize some metric of the phenomenon in question and thus might use the particular measurement standards directly when measuring actions or outcomes, or indirectly by using the resulting ratings and rankings as criteria in subsequent assessments or decisions about actions. In some respects all actual and potential users may be considered a target of indicators in that these explicit, often technical norms seek to affect the way in which measurements are taken, or at least oblige those who might measure the same phenomenon differently to justify their deviant metric through comparison.¹³ I distinguish, however, within the broader group of the “users” the “targets of the rules.” These “targets” comprise the subset of users who either are themselves being measured or who act and speak on behalf of the objects or institutions being measured. Although indicators are not always consciously intended to affect the behavior of the measured, I posit the measured as the ultimate target of any indicator.¹⁴

¹² Note that this *political* notion of demand differs from the standard consumption-base notion of demand in economics.

¹³ In such comparisons, there often is some deference to the measurement standards that were prior (if widely used). For an analysis of why divergent metrics might not converge even when the competing suppliers appear to have nearly identical goals and the divergence is costly to all of them, see Claire Lauterbach’s analysis of the competing body counts for civilian casualties in the Iraq war of 2003/04: Claire Lauterbach, “The Costs of Cooperation: Civilian Casualty Counts in Iraq,” *International Studies Perspectives* 8:4 (November 2007): 429–45.

¹⁴ Prompting a behavioral adaptation might in fact undermine the usefulness of the indicator as a scientific measure. Indicators might of course measure characteristic rather than behavior, or measure aspects of large organizations or institutions. In such cases the behavioral response might come from senior decision-makers within those organizations or other actors who might be able to bring about changes in the observed characteristics or in observability.

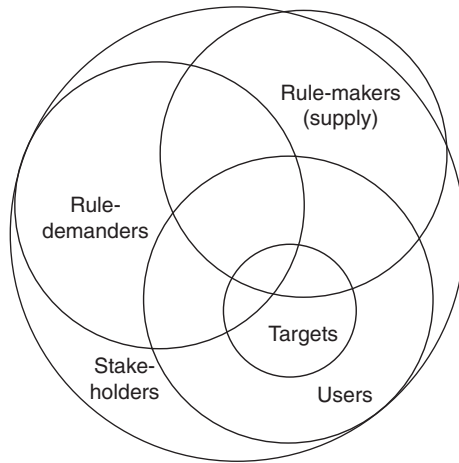


Fig. 2.1 Multi-stakeholder Conceptual Model

Source: Tim Büthe, 2012.

Any of these particular sets of stakeholders may more or less overlap (see Figure 2.1), with important implications for the relationships of power among the stakeholders. Moreover, there are good reasons to expect that, for many indicators, the group of stakeholders is even broader than the combination of rule-demanders, rule-makers, and users (including targets). Adam Smith famously remarked that “people of the same trade seldom meet...[without that] the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.” When economic or other social activities are measured and thus regulated, parties beyond the subsets of stakeholders identified above are likely to benefit or be negatively affected, even when there is no ill intent or violation of antitrust laws as suggested by Smith’s warning, because such activities often have externalities.

Recognizing the possible divergence among the various stakeholders who are affected by a particular measurement standard (or its absence), especially the divergence between the rule-makers and the actors who use those rules (including the “target” subset), has implications for both positive analysis and normative assessment. For positive analysis, it underscores the need to examine—separately—demand, supply, and use, as well as the behavioral adaptation by those whom I label the ultimate “targets” of indicators. I will offer such an analysis in the next section.

Figure 2.1 can also be useful for normative analysis. Ideal-typical “self-regulation,” for instance, entails complete overlap of all three groups. The greater the extent to which the rule-makers and the targets of the rules do *not* coincide (assuming that preferences meaningfully diverge), the more does the supply of indicators entail the exercise of power. Specifically, power will matter more, the more the targets of the rules are excluded from rule-making or (looking at it from

the other direction) the more the set of rule-makers extends beyond the users and especially the targets of the rules. And regardless of whether “third parties” benefit from the private rules or are negatively affected by them, the extent to which such passive stakeholders are excluded from the rule-making process determines how much governance by indicators diverges from the liberal notion of “democracy,” in which all who are affected by a decision have a voice in the decision-making.¹⁵

From conceptual model toward explanation

In this section, I offer a sketch of possible explanations for why rule-demanders might seek indicators, why rule-makers might supply them, why users might use them—and why the targets of these measurement standards tend to respond in ways that make indicators a power resource in the Dahlian sense of allowing one actor to cause another actor to do something the latter would not otherwise have done.¹⁶ I do not claim to provide here a comprehensive explanation of why certain indicators exist and others do not, nor of the precise effects of particular indicators; rather, I seek to illustrate the usefulness of the conceptual model as a framework for such analyses. Most of the reasons discussed below were deductively derived from thinking about the regulation of internationally integrated markets, but I submit that they are also applicable to other kinds of rules and rule-making. To underscore the applicability of the general reasoning to indicators (understood as measurement standards), I suggest how it may be applied to several of the specific indicators that are examined in detail in later chapters of this book.

Demand

Demand for rules can be simply a functional pursuit of Pareto-improving efficiency gains. At the micro level, within firms, for instance, the demand for rules that govern the internal gathering and exchange of information is often motivated by wanting to lower transaction costs or increase reliability.¹⁷ At the macro level, the demand for market rules (such as measurement standards) might be expected from actual or potential market participants seeking to facilitate commercial transactions, or from political leaders seeking to increase the efficient operation of markets in implicit or explicit recognition of the fact that markets do not spontaneously arise

¹⁵ See, e.g., Robert A. Dahl, *A Preface to Economic Democracy* (Berkeley: University of California Press, 1985); Eyal Benvenisti and George Downs, “Toward Global Checks and Balances,” *Constitutional Political Economy* 20:3/4 (September 2009): 366–87.

¹⁶ Robert A. Dahl, “The Concept of Power,” *Behavioral Science* 2:3 (July 1957): 201–15.

¹⁷ Such benefits have been amply documented in micro-economic studies, e.g., DIN (Deutsches Institut für Normung), *Gesamtwirtschaftlicher Nutzen der Normung, Zusammenfassung der Ergebnisse: Wissenschaftlicher Endbericht mit Praktischen Beispielen* (Berlin: Beuth Verlag, 2000), esp. 14f; Gerald C. Joynt, “Standards Pay Dividends,” in Rowen Glie (ed.), *Speaking of Standards* (Boston: Cahner Books, 1972), 182–93.

but require rules well beyond a guarantee of private property rights.¹⁸ And at the international level, the push for the harmonization of financial reporting rules, for instance, was motivated in large part by political leaders' expectation that common rules for measuring and reporting financial assets, liabilities, expenditures, etc. would lead to a more efficient allocation of capital.¹⁹

Demand for an indicator need not be driven by commercial interests, nor entail a specific demand for a particular measure, but may be prompted by a sense that some members of the global community have information needs, combined with an entrepreneurial idea for how to meet those needs. *US News & World Report*, as Espeland and Sauder note, initially vastly underestimated the potential demand for college and law school rankings. Magazine editors merely conjectured that, among the ideas prompted by the new branding motto for the magazine under its new owner in the 1980s ("News you can use"), this was one for which there would be enough demand to sell additional magazines.²⁰ Similarly, in the non-commercial realm, the idea for the US Fund for Peace's Failed States Index arose, according to Nehal Bhuta's account, out of a sense among Fund staff that it would be desirable (i) to have some systematic way of identifying countries whose precarious domestic political situation warranted additional, targeted aid for political stabilization from Western governments; and (ii) to have better means of identifying internal security threats in advance of the outbreak of violence.²¹

The desire to have seemingly objective, scientific measures of performance or outcomes to use in naming and shaming those with poor performance provides a third possible, more overtly political explanation for demanding an indicator. The demand for many corporate social responsibility (CSR) standards may be explained in this way; indicators of human rights laws, policies, or outcomes often started out with such a political use already intended.²² And Freedom House's indices of

¹⁸ E.g., Neil Fligstein, *The Architecture of Markets: An Economic Sociology of Twenty-First-Century Capitalist Societies* (Princeton: Princeton University Press, 2001); Steven K. Vogel, *Freer Markets, More Rules: Regulatory Reform in Advanced Industrial Countries* (Ithaca, NY: Cornell University Press, 1996).

¹⁹ Büthe and Mattli, *New Global Rulers* (n. 5); Kees Camfferman and Stephen A. Zeff, *Financial Reporting and Global Capital Markets: A History of the International Accounting Standards Committee, 1973–2000* (Oxford: Oxford University Press, 2000); Leonardo Martinez-Diaz, "Strategic Experts and Improvising Regulators: Explaining the IASC's Rise to Global Influence 1973–2001," *Business and Politics* 7:3 (December 2006). Demand for indicators that are fully explained by the affected private parties' correctly anticipated efficiency gains raises few political issues, though to the extent that articulating such demands is costly, anticipated gains might still fail to result in observable demands due to collective action problems.

²⁰ Wendy Nelson Espeland and Michael Sauder, "The Dynamism of Indicators" in Kevin E. Davis et al. (eds), *Governance by Indicators: Global Power through Quantification and Rankings* (New York: Oxford University Press, 2012), 86–109, esp. 88f.

²¹ Nehal Bhuta, "Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order," in Davis et al. (eds), *Governance by Indicators* (2012) *ibid*, 132–62.

²² On CSR standards, see David Vogel, *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility* (Washington, DC: Brookings Institution Press, 2005) and "Private Global Business Regulation," *Annual Review of Political Science* 11 (2008): 261–82; on human rights indicators, see the chapters by Rosga and Satterthwaite, Gallagher and Chuang, Zaloznaya and Hagan, and Satterthwaite in this volume.

political rights and civil liberties—and the resulting Freedom in the World ranking—were, according to their own official history, explicitly intended from the beginning in the early 1970s to advance FH’s mission “to expand freedom around the world.”²³ Members of the US-based organization who proposed developing such measures sought not only to show that citizens of Western democracies actually (rather than just nominally/legally) enjoyed superior political rights and civil liberties vis-à-vis citizens of the Communist world, but also to nudge governments in the rest of the world toward improving the fairness of elections, ease restrictions on the formation of political or civil society organizations, and ensure free speech without retribution—to name just some of the key criteria for “good” FH scores.²⁴

Strikingly, many indicators—including the majority of those discussed in the chapters of this volume—were developed and are maintained by private (non-governmental) bodies. This raises the additional question: “Why *private* indicators?” Why might anyone seek private rather than public rules? One possible reason is simply that the non-governmental realm is where the pertinent expertise is. This is especially likely when the phenomenon to be measured is changing at a fast pace, it becomes challenging for anyone who is removed from the scientific and technological frontier—such as a regulatory generalist in a government agency—to have up-to-date information about what exactly might need to be measured and how it can be done. Abbott and Snidal’s assessment of the limitations of governments in the regulation of production also applies more generally: as a consequence of the “scale and structure of contemporary global production . . . , no actor group, even the advanced democratic state, possesses all the competencies required for effective regulation.”²⁵ Such reduced capacity of governments to provide measurement standards without incurring enormous expenses might prompt governments or members of civil society to seek private rules instead.²⁶

The literature on the delegation of public authority also identifies more overtly political motivations for delegation to a private rule-maker, such as shifting

²³ *Freedom House: A History*, available at <<http://www.freedomhouse.org/content/our-history>> (last accessed April 15, 2012).

²⁴ See, e.g., Joseph E. Ryan, “Survey Methodology,” *Freedom Review* 25:1 (February 1994): 9–22.

²⁵ Kenneth W. Abbott and Duncan Snidal, “The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State,” in Walter Mattli and Ngaire Woods (eds), *The Politics of Global Regulation* (Princeton: Princeton University Press, 2009), 44–88, esp. 44, 68.

²⁶ Government deference to private bodies may be motivated by a desire to achieve efficiency gains from specialization and to benefit from specific expertise that already exists and is more efficiently maintained in the private sector instead of having the government acquire/maintain it “in house.” It may, however, also be motivated by a lack of political will. Even when such a demand for private rules entails explicit delegation of governmental authority to private bodies, it need not entail complete abdication by governments. In Europe, product safety regulation, for instance, has increasingly entailed governments laying down general principles but leaving it to private bodies at the regional level, such as CEN and CENELEC for general product and electrotechnical standards, to specify the best way to meet those broad public policy objectives without introducing non-tariff barriers to trade or otherwise fragmenting the Common Market through cross-national divergence in regulatory measures. See Michelle Egan, *Constructing a European Market: Standards, Regulation, and Governance* (New York: Oxford University Press, 2001); Harm Schepel, *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets* (Portland: Hart Publishers, 2005).

responsibility in order to seemingly de-politicize the issue and avoid subsequent blame.²⁷ Another politically important motivation for delegation is to institutionalize a policy bias favored by the current, temporary political majorities.²⁸ Bias in favor of the private sector, for instance, may be effectively institutionalized by delegating rule-making to a private body in which “members” pay for a seat at the table, thus reducing the involvement of non-commercial civil society interests.²⁹

Individuals and civil society groups might, of course, also seek private rule-making. In fact, many of the most prominent private indicators in the international political economy were prompted by demands from societal actors who felt that their *non-material* interests were affected by private commercial transactions between other private parties, and that existing or missing public rules failed to safeguard their interests sufficiently. Social activists have thus demanded measurement standards for forestry practices,³⁰ environmental management,³¹ corporate social responsibility,³² and many other aspects of the behavior of key actors in the global economy. Often these demands were initially aimed at governments but were transformed into demands for private rules when governments proved unresponsive or, in the case of rules for global markets, when inter-governmental negotiations were excruciatingly slow.³³

Supply

Functional explanations of the demand for private regulation, which emphasize anticipated efficiency gains, may be read to imply that such efficiency gains will also

²⁷ Tim Büthe, *The Dynamics of Principals and Agents: Institutional Persistence and Change in U.S. Financial Regulation, 1934–2003*, unpublished manuscript, Duke University, 2006–2010; Morris P. Fiorina, “Legislative Choice of Regulatory Forms: Legal Process or Administrative Process?,” *Public Choice* 39:2 (September 1982): 33–66.

²⁸ Terry M. Moe, “Power and Political Institutions,” *Perspectives on Politics* 3:2 (June 2005): 215–33.

²⁹ Tim Büthe, “The Politics of Food Safety in the Age of Global Trade: The Codex Alimentarius Commission in the SPS-Agreement of the WTO,” in Cary Coglianese, Adam Finkel, and David Zaring (eds), *Import Safety: Regulatory Governance in the Global Economy* (Philadelphia: University of Pennsylvania Press, 2009), 88–109; Büthe and Mattli, *New Global Rulers* (n. 5), ch. 9.

³⁰ Tim Bartley, “Certifying Forests and Factories: States, Social Movements, and the Rise of Private Regulation in the Apparel and Forest Products Field,” *Politics & Society* 31:3 (September 2003), 433–64; Benjamin Cashore, Graeme Auld, and Deanna Newsom, *Governing Through Markets: Forest Certification and the Emergence of Non-State Authority* (New Haven: Yale University Press, 2004); Errol Meidinger, Chris Elliott, and Gerhard Oesten (eds), *Social and Political Dimensions of Forest Certification* (Remagen-Oberwinter: Forstbuch Verlag, 2003).

³¹ Magali A. Delmas, “The Diffusion of Environmental Management Standards in Europe and in the United States: An Institutional Perspective,” *Policy Sciences* 35:1 (2002): 91–119; Aseem Prakash and Matthew Potoski, *The Voluntary Environmentalists* (New York: Cambridge University Press, 2006).

³² D. Vogel, *The Market for Virtue* (n. 22).

³³ In some cases, these demands from societal stakeholders prompted formal government delegation to a private standard-setter, because it allowed governments to achieve a policy outcome favored by its domestic constituents but without running afoul of the government’s obligations under international law, such as linking market access to labor conditions or environmental performance of the producer, which is impermissible under WTO rules (see Bartley, “Certifying Forests and Factories” (n. 30); D. Vogel, “Private Global Business Regulation” (n. 22)).

induce potential private-sector rule-makers to supply those rules. Such an ideal overlap of demanders and suppliers, however, rarely just happens but may need to be politically established, since collective action problems and especially incentives to free-ride can be expected to undermine the functionalist supply of private regulation. This logic is nicely illustrated by the IRIS/GIIRS indicators for impact investing examined in Sarah Dadush's chapter. The interest among investors in investment opportunities that offer not just a financial return but simultaneously a chance to have a "positive impact" created a multi-stakeholder demand for metrics to communicate key features of such investment opportunities, such as the magnitude and likelihood of the "positive impact" in credible and comparable ways.³⁴ The supply of such measurement standards promised to literally create a market, bringing not just a marginal Pareto-improvement but substantial benefits for many stakeholders. Yet, the small size of initial investments of this kind and the risk that multiple, uncoordinated metrics would undercut the market, led to under- or more precisely non-supply until the Rockefeller Foundation, JP Morgan, and US AID joined forces and launched the Global Impact Investing Network to overcome collective action and coordination problems among the stakeholders to develop the IRIS/GIIRS indicators.

It is also possible that some individuals or groups with the requisite technical expertise might supply an indicator out of altruism. Although some argue that even altruism is inherently impure because it benefits the ostensible altruist, at a minimum, through an increased sense of self-worth,³⁵ such altruists might have no prior (and in the purest form of altruism maybe also no subsequent) stake in the phenomenon or the specific measure adopted. Hence, the circle depicting the set of suppliers in Figure 2.1 extends slightly beyond the total set of stakeholders.

Beyond situations where altruism explains the supply of indicators, they are likely to be undersupplied—given the cost of developing high-quality measurement standards—unless rule-making also brings some economic, political, or other private benefits for those who participate or contribute to it: material gains, the institutionalization of disproportionate influence over the content of the rules, or lock-in of a particular policy. I briefly discuss each of these rationales in turn.

Socio-political actors may gain material benefits from supplying rules either directly or indirectly through the impact of the indicator on the structure of the relevant market, which may make demand and supply of such an indicator mutually reinforcing. Direct material gain as the motivation to supply rules is clearly apparent in Espeland and Sauder's analysis of *US News and World Report's* law school rankings. The expectation of increased magazine sales motivated the magazine's editors not just to "demand" but also to supply this indicator—in order to capture the expected material gains that would accrue directly to the rule-maker.³⁶ Related but analytically distinct is the willingness of commercial actors to cooperate in (or contribute to)

³⁴ Sarah Dadush, "Impact Investment Indicators: A Critical Assessment," this volume, pp. 392–434.

³⁵ James Andreoni, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal* 100:401 (June 1990): 464–77.

³⁶ Espeland and Sauder, "The Dynamism of Indicators" (n. 20).

rule-making because it allows them to create or retain barriers to entry, or create or reinforce an oligopolistic market structure, guaranteeing profits greater than the cost of rule-making.³⁷ Peters et al. argue, for instance, that this explains the supply of bond ratings.³⁸

Supplying an indicator can also bring political gains by ensuring immediate influence over its content and construction and often by institutionalizing and thus safeguarding the rule-making privilege of a particular set of experts (guaranteeing greater influence for a particular set of interests vis-à-vis other stakeholders).³⁹ The process of setting such standards may, of course, be specifically designed to be inclusive of a broad range of stakeholders, as Christopher Stone illustrates in his chapter on public safety and justice, but such inclusiveness is rare in global governance.⁴⁰ The supply of indicators as a means of ensuring influence is particularly attractive if there is a real prospect that an alternative indicator may otherwise be supplied by an actor with differing preferences or in a body with fewer opportunities to exert influence over the content of the rules. The threat of government regulation, for instance, often provides a powerful incentive for the supply of market regulations by the private sector.⁴¹ In the realm of international finance, Eric Helleiner argues that the consortium of major international banks known as the Institute of International Finance stepped in to develop standards for sovereign debt restructuring above all to break the momentum of an inter-governmental initiative for an IMF-sponsored international bankruptcy mechanism preferred by, and likely more favorable to, debtor countries.⁴²

At the same time, such distributional implications of the supply of indicators create incentives for establishing competing rule-makers. Such competition between multiple

³⁷ A. Claire Cutler, Virginia Haufler, and Tony Porter, "Private Authority and International Affairs," in Cutler, Haufler, and Porter (eds), *Private Authority and International Affairs* (Albany: State University of New York Press, 1999), 3–28, esp. 7; Matthew Potoski and Aseem Prakash (eds), *Voluntary Programs: A Club Theory Approach* (Cambridge, MA: MIT Press, 2009).

³⁸ Anne Peters, Lucy Koechlin, and Greta Fenner Zinkernagel, "Non-State Actors as Standard Setters: Framing the Issue in an Interdisciplinary Fashion," in Peters et al. (eds), *Non-State Actors as Standard Setters* (Cambridge: Cambridge University Press, 2009), 1–32, esp. 2f.

³⁹ The result can be less stringent rules with more loopholes than would have come about through a more inclusive process (see Büthe, *The Dynamics of Principals and Agents* (n. 27), but the opposite is also possible: firms that already use clean technology or produce products with multiple safety features, for instance, may seek to establish more stringent standards of environmental or consumer protection in order to gain an advantage vis-à-vis their competitors (see David Vogel, *Trading Up: Consumer and Environmental Regulation in a Global Economy* (Cambridge, MA: Harvard University Press, 1995)). Prakash and Potoski document such a competitive dynamic for voluntary certifications for environmental impact assessment, especially ISO 14000-series standards: Aseem Prakash and Matthew Potoski, "Racing to the Bottom? Trade, Environmental Governance, and ISO 14001," *American Journal of Political Science* 50:2 (April 2006): 350–64.

⁴⁰ Christopher Stone, "Problems of Power in the Design of Indicators of Safety and Justice in the Global South," in this volume, pp. 281–94.

⁴¹ E.g., Virginia Haufler, *The Public Role of the Private Sector: Industry Self-Regulation in a Global Economy* (Washington, DC: Carnegie Endowment for International Peace, 2001); Büthe and Mattli, *New Global Rulers* (n. 5).

⁴² Eric Helleiner, "Filling a Hole in Global Financial Governance? The Politics of Regulating Sovereign Debt Restructuring," in Mattli and Woods (eds), *The Politics of Global Regulation* (2009), 89–120.

alternative metrics is in fact quite common, for instance in the realm of “fair trade,” where indicators developed by NGOs often compete against indicators developed by firms,⁴³ but also, for example, among indicators for failed or “fragile” states.⁴⁴

Use of indicators

Even when measurement standards are written by someone with no power over potential users, governments (or others in a position of authority) may subsequently require their use, resulting in widespread adoption. Domestically in the United States, for example, the explicit delegation of health and safety standard-setting to private bodies has often been accompanied by federal and state-level laws and regulations mandating their use, and even compliance.⁴⁵ In more recent years, this practice has also become common inter- and transnationally. The Canadian Medical Devices Regulation SOR/98-282, for instance, requires manufacturers that sell medical devices in Canada to use a measurement standard developed by the non-governmental International Organization for Standardization (ISO 13485) to assess their own quality management systems and report the results.⁴⁶ Similarly, the measurement standards used by the Programme for International Student Assessment were largely developed as explicit technical norms by a transnational body, the International Association for the Evaluation of Educational Achievement. After some internal conflicts discussed by von Bogdandy and Goldmann, the member states of the OECD jointly adopted them (with some modifications) and each of them mandated their use in the domestic PISA tests.⁴⁷

Notwithstanding the importance of public measures that make the use of a particular metric mandatory, many indicators are widely used even when not

⁴³ E.g., Margaret Levi and April Linton, “Fair Trade: A Cup at a Time?” *Politics & Society* 31:3 (September 2003): 407–32; Daniel Jafee, *Brewing Justice: Fair Trade Coffee, Sustainability, and Survival* (Berkeley: University of California Press, 2007); Marie-Christine Renard, “Fair Trade: Quality, Market and Conventions,” *Journal of Rural Studies* 19:1 (January 2003): 87–96; Peter L. Taylor, “In the Market But Not of It: Fair Trade Coffee and Forest Stewardship Council Certification as Market-Based Social Change,” *World Development* 33:1 (January 2005): 129–47; Laura T. Reynolds, Douglas L. Murray, and John Wilkinson (eds), *Fair Trade: The Challenges of Transforming Globalization* (London–New York: Routledge, 2007).

⁴⁴ Bhuta, “Governmentalizing Sovereignty” (n. 21).

⁴⁵ Ross E. Cheit, *Setting Safety Standards: Regulation in the Public and Private Sectors* (Berkeley: University of California Press, 1990); Robert W. Hamilton, “The Role of Nongovernmental Standards in the Development of Mandatory Federal Standards Affecting Safety or Health,” *Texas Law Review* 56:8 (November 1978): 1329–484; Stewart Macaulay, “Private Government,” in Leon Lipson and Stanton Wheeler (eds), *Law and the Social Sciences* (New York: Russell Sage Foundation, 1986), 445–518; Liora Salter, *Mandated Science: Science and Scientists in the Making of Standards* (Dordrecht: Kluwer Academic Publishers, 1988).

⁴⁶ ISO/IEC, *Using and Referencing ISO and IEC Standards for Technical Regulations* (Geneva: International Organization for Standardization and International Electrotechnical Commission, 2007), 23. On the ISO, see Craig N. Murphy and JoAnne Yates, *The International Organization for Standardization: Global Governance through Voluntary Consensus* (London–New York: Routledge, 2008); Büthe and Mattli, *New Global Rulers* (n. 5), esp. chs 6–7.

⁴⁷ Armin von Bogdandy and Matthias Goldmann, “Taming and Framing Indicators: A Legal Reconstruction of the OECD’s Programme for International Student Assessment (PISA),” in this volume, pp. 52–85, esp. 54–62.

required. They may even be used by socio-political actors who would have preferred a world in which no such indicator existed. To understand why, it is useful to distinguish between several possible reasons, starting again with strictly Pareto-improving economic incentives (which may nonetheless have distributional implications), then moving to more overtly political incentives.

In any context where there are infrequent arm's-length interactions between two or more parties, information asymmetry between them can impede collaboration and exchanges that would benefit both sides. This finding has been well established for product markets where, as George Akerlof famously pointed out, information asymmetry between buyers and sellers depresses quality and size of markets,⁴⁸ but the logic also applies more broadly. Both sides here have economic incentives to seek to overcome or at least reduce the information asymmetry. Standards for how to measure and report quality can help overcome the problem, at least if there is monitoring and/or enforcement, for instance through credible third-party certification or efficient adjudication. Rankings and ratings that are regularly updated, based on de novo application of the same measurement standards by a third party (i.e., someone who has no stake in the distribution of benefits between the parties), provide such monitoring. Their use might then simply be explained by the informational benefits they promise. Multinational corporations that regularly buy the latest indicators of "rule of law" or other political risks discussed by Katharina Pistor presumably are motivated at least in part by an anticipation of such informational benefits.⁴⁹ Employers similarly may look to rankings of universities as an information shortcut regarding the quality of the training received (or the likely average quality of a student admitted there), especially in the case of universities about which the employer has limited local/personal knowledge, that is, when the condition of infrequent, arm's-length relationships holds.⁵⁰ Those who are to be measured may even invite being measured when they expect "consumer demand" to exist for certain information, so that the existence of indicators would provide opportunities for product differentiation and consequently more effective marketing, higher prices, etc.—i.e., use may feed back into demand.⁵¹ A similar reasoning appears to

⁴⁸ George A. Akerlof, "The Market for Lemons: Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84:3 (August 1970): 488–500.

⁴⁹ Katharina Pistor, "Re-Construction of Private Indicators for Public Purposes," in this volume, pp. 165–79, esp. 167.

⁵⁰ See Espeland and Sauder, "The Dynamism of Indicators" (n. 20).

⁵¹ The political economy literature tends to show that those with a valuable "brand" name have a stronger incentive to provide accurate information on their own, but third-party monitoring and enforcement can still make a substantial difference. See Dennis G. Arnold, "Philosophical Foundations: Moral Reasoning, Human Rights, and Global Labor Practices," in Laura Pincus Hartman (ed.), *Rising Above Sweatshops: Innovative Approaches to Global Labor Challenges* (Santa Barbara, CA: Greenwood Publishing, 2003), 77–99; Tara J. Radin, "Chiquita Brands International Inc.: Values-Based Management and Corporate Responsibility in Latin America," in Laura Pincus Hartman (ed.), *Rising Above Sweatshops: Innovative Approaches to Global Labor Challenges* (Santa Barbara, CA: Greenwood Publishing, 2003), 353–83; Tim Bartley, "Corporate Accountability and the Privatization of Labor Standards: Struggles over Codes of Conduct in the Apparel Industry," *Research in Political Sociology* (2005), 211–44; Richard M. Locke, Fei Qin, and Alberto Brause, "Does Monitoring Improve Labor Standards? Lessons from Nike," *Industrial and Labor Relations Review* 61:1 (October 2007): 3–31.

have motivated the European business schools that, as Espeland and Sauder report, pushed for an international ranking of MBA programs by the *Financial Times*, which they expected to give greater weight to what they saw as the underappreciated strengths of a European approach to management education.⁵²

Network externalities can also create incentives for the use of indicators.⁵³ Network externalities arise whenever the benefit one derives from a particular choice increases with the number of others who have made the same choice, or when the value of a product increases with the extent to which complementary products are available. Measuring physical objects using the international system of units may be attractive because the clearly specified relationships between them facilitates calculation, but it becomes greatly more attractive the larger the number of others who use these units, because it increases the availability and lowers the price of instruments for measuring in those units and it increases the number of people who will correctly understand the resulting information. These benefits of the size of the “network” are positive externalities to the extent that they are not reflected in the cost of producing nor the price (if any) paid for using an indicator. Such positive externalities create an incentive to use the indicator.⁵⁴ In fact, they provide an incentive to use a particular indicator even if the user considers the indicator poorly designed and its use objectionable. Many prospective US law students, for example, may believe that the *US News* indicator assigns too much weight to factors they consider irrelevant or even that the prominence of this indicator is detrimental to the legal education that they personally seek, but most nonetheless consider *US News*’ ranking of law schools when deciding between competing offers of admission because they know prospective employers will do so.⁵⁵

⁵² Espeland and Sauder, “The Dynamism of Indicators” (n. 20), p. 106, based on Linda Wedlin, *Ranking Business Schools: Forming Fields, Identities and Boundaries in International Management Education* (Cheltenham, UK: Edward Elgar).

⁵³ As generally for measurement and product standards, see Paul A. David, “Clio and the Economics of QWERTY,” *American Economic Review* 75:2 (Papers and Proceedings, May 1985): 332–7; Michael L. Katz and Carl Shapiro, “Network Externalities, Competition, and Compatibility,” *American Economic Review* 75:3 (June 1985): 424–40.

⁵⁴ Some argue that such network externalities are quite rare (see Stan J. Liebowitz and Stephen E. Margolis, “Network Externality: An Uncommon Tragedy,” *Journal of Economic Perspectives* 8:2 (Spring 1994): 133–50). Yet, they appear to be very common in the realm of information and communication technology—the value of having a cell phone, fax machine, or network-able computer, for instance, is largely a function of the number of others with whom I can communicate using this technology (see Joseph Farrell and Garth Saloner, “Competition, Compatibility and Standards: The Economics of Horses, Penguins and Lemmings,” in H. Landis Gabel (ed.), *Product Standardization and Competitive Strategy* (Elsevier Science Publishers, 1987), 1–21)—and many find them to be generally pervasive. See, e.g., Jean Tirole, *The Theory of Industrial Organization* (Cambridge, MA: MIT Press, 1988); Oz Shy, *The Economics of Network Industries* (Cambridge: Cambridge University Press, 2001); David Singh Grewal, *Network Power: The Social Dynamics of Globalization* (New Haven: Yale University Press, 2008).

⁵⁵ See Michael Sauder and Ryon Lancaster, “Do Rankings Matter? The Effects of U.S. News and World Report Rankings on the Admission Process of Law Schools,” *Law & Society Review* 40:1 (March 2006): 105–34. Similar concerns have been raised regarding college rankings: David S. Webster, “Rankings of Undergraduate Education in US News & World Report and Money: Are They Any Good?,” *Change* 24:2 (March/April 1992): 18–31.

A more overtly political reason for the use of an indicator is the demand by some subset of stakeholders for systematic information. Such a desire may be articulated by the original “demanders” and thus be directly tied to a specific measurement standard: direct pressure from activist NGOs (or consumer demand stimulated by such civil society groups) has led many businesses to commit at least rhetorically to various “fair trade” and other corporate social responsibility standards and to report their performance on the measures specified by those standards.⁵⁶

Alternatively, the desire for systematic information may originate with individuals or groups who have no relationship to those whose demand and supply brings the indicator into existence in the first place. In fact, the initial demand for systematic information on phenomenon X may be articulated by stakeholders who have no knowledge of the particular standard, but its availability and possible usefulness for providing a measure of X then may ensure its widespread use, especially when the collective action problems noted above impede the supply of alternative, better indicators. Thus, the immunization coverage indicators analyzed by Angelina Fisher were initially designed to provide narrow measures of the effectiveness of very specific immunization programs which sought to achieve blanket coverage against a small number of diseases. Yet policymakers, commercial actors, and scholars also sought more general measures of health status and health care quality or coverage, especially in the weakly institutionalized environments of developing countries for which reliable, comparable statistics are often scarce.⁵⁷ The ready availability of the immunization indicators then led to their re-interpretation and adoption as measures of any number of health policy and health outcomes.⁵⁸

The ultimate targets: Behavioral change by the measured

Research in psychology and behavioral economics has shown that knowingly being observed leads human beings to change their behavior in ways that increase compliance with broader social norms or, more specifically, change their behavior in ways that the observed expects to elicit greater approval from the observer. And

⁵⁶ For discussion of the fair trade example, see Levi and Linton, “Fair Trade: A Cup at a Time?” (n. 43); Cashore, Auld, and Newsom, *Governing Through Markets* (n. 30); John J. Kirton and Michael J. Trebilcock, *Hard Choices, Soft Law: Voluntary Standards in Global Trade, Environment and Social Governance* (Burlington, VT: Ashgate, 2004); Jaffee, *Brewing Justice* (n. 43); Reynolds, Murray, and Wilkinson (eds), *Fair Trade* (n. 43). The threat of government regulation, noted above as a possible reason for the private supply of indicators may also explain their use if there is a strong expectation that such use can forestall public rules among actors who can overcome the collective action problems.

⁵⁷ On the diversity of state capacity and strength of political institutions across developing countries, see, e.g., Steven Levitsky and María Victoria Murillo, “Variation in Institutional Strength,” *Annual Review of Political Science* 12 (2009): 115–33.

⁵⁸ Angelina Fisher, “From Diagnosing Under-Immunization to Evaluating Health Care Systems: Immunization Coverage Indicators as a Technology of Global Governance,” this volume, pp. 217–46.

indicators communicate the desired behavior (or more precisely what the observer would like to see) subtly but very clearly through scores and rankings, making them ideal instruments for eliciting behavioral changes through observation.

Beyond this apparently inherent human tendency to want to gain the approval of those observing us, which is internal to the measured, the ultimate targets of an indicator face both material and socio-political incentives for behavioral adaption—probably, though not necessarily, toward the ideal defined by the indicator. The economic (materialistic) incentives may, as before, operate directly or indirectly. As Nikhil Dutta points out, for instance, Freedom House's ratings of a country's political rights and civil liberties are used by the Millennium Challenge Corporation in making decisions about the disbursement of MCC-administered US foreign aid.⁵⁹ Such direct economic incentives are likely to elicit attempts not only to maintain a good score but to "improve" one's score. And seeking an improvement in its score, a country would generally be expected to make changes in policies or political institutions that bring the country closer to the FH ideal. Note, however, that this need not be the only way to improve one's score: falsifying information and generally making objectionable behavior harder to observe might be the preferred (and possibly in the short run the more efficient) way.

Law school rankings illustrate a slightly more indirect operation of material incentives. Law school deans and faculty may mostly be opposed to the use of this indicator as a measure of the quality of legal education, but they overwhelmingly feel compelled to pay close attention and seek to improve their schools' rank, given that prospective students and future employers of those potential students are using this indicator. And political-legal incentives may reinforce more indirect material incentives: indicators effectively (seek to) define "best practice" and not living up to it can be costly. A firm that does not implement widely accepted standards for workplace safety, for instance, may face a higher insurance premium.⁶⁰ Should an accident occur and lead to a lawsuit, having achieved the highest rating does not necessarily safeguard the firm against losing in court. But, as Fabrizio Cafaggi has found, *not* having implemented what is widely considered "best practice" will in many jurisdictions substantially increase the risk of being found negligent.⁶¹

A further important socio-political incentive may arise from overt peer pressure. In what is considered an exemplary case of pro-active industry self-regulation after the Three-Mile Island nuclear power plant accident, the Institute of Nuclear Power Operations created a "new responsibility-centered industrial culture."⁶² Yet, this "success," Bridget Hutter points out, was only achieved through sustained peer pressure and mutual assurances among the firms in that industry to avoid

⁵⁹ Nikhil K. Dutta, "Accountability in the Generation of Governance Indicators," this volume, pp. 437–64, esp. 441.

⁶⁰ Richard V. Ericson, Aaron Doyle, and Dean Barry, *Insurance as Governance* (Toronto: University of Toronto Press, 2003).

⁶¹ Fabrizio Cafaggi, "Private Regulation in European Private Law," EUI Law Working Paper no. 2009/31.

⁶² Joseph Rees, *Hostages of Each Other* (Chicago: University of Chicago Press, 1994).

government regulation.⁶³ Indicators—and the resulting rating or ranking—should make it easier for interested parties within the social group or even beyond (such as from the “international community”) to bring pressure on the targets to “improve” their behavior in such a way that it yields a higher score (i.e., a higher rating or maybe even a higher ranking).

Applications of the conceptual model

To illustrate the analytical usefulness of the conceptual model, I apply the four-stakeholder model to two specific examples of influential indicators examined in greater detail in other chapters in this volume: the ranking of US law schools conducted annually since 1987 by the popular magazine *US News and World Report* (*USN*), and the Freedom in the World rating, published since 1973 by the heavily government-funded but otherwise non-governmental organization Freedom House. Applying the conceptual model to these specific cases and re-drawing the generic Figure 2.1 with proper names allows me to highlight relationships of power among the key stakeholder groups, including important changes over time.

US News & World Report's law school rankings

USN ranks the reputation (“quality assessment”), selectivity, placement record, and faculty resources for each law school that is fully accredited by the American Bar Association. The “quality assessment” score is an inevitably subjective measure of each school’s reputation, based on surveys among legal academics, practitioners, and judges. The reputation component is given a nominal weight of 40 percent in determining each law school’s overall score.⁶⁴ The remainder is mostly comprised of ostensibly objective components for selectivity (acceptance rate and the entering class’s median undergraduate grades and median scores on the national law school “aptitude” test, LSAT) and for placement success (employment rates for graduates at graduation and after nine months, plus the passage rate of first-time takers of the tests for admission to the bar association in the US state where the school is located). Measures of “faculty resources” (expenditure per student for instruction,

⁶³ Bridget M. Hutter, “The Role of Non-State Actors in Regulation,” in Gunnar Folke Schuppert (ed), *Global Governance and the Role of Non-State Actors* (Baden-Baden: Nomos Verlag, 2006), 63–79.

⁶⁴ 40 percent is the “nominal weight” in that each component’s actual influence on the overall score is also a function of the components’ variability; see, e.g., Sauder and Lancaster, “Do Rankings Matter?” (n. 55): 112–17. In the initial 1987 ranking, only a survey of law school deans was conducted. Later, the academic sample was enlarged to also include three members of each school’s faculty; then a survey of legal practitioners (especially hiring partners, senior officials in state attorney generals’ offices, and judges) was added. For a detailed discussion of the reputational component of the *USN* ranking, see Richard Schmalbeck, “The Durability of Law School Reputation,” *Journal of Legal Education* 48:4 (December 1998): 568–90 and Jeffrey Evans Stake, “The Interplay Between Law School Rankings, Reputations, and Resource Allocations: Ways Rankings Mislead,” *Indiana Law Journal* 82:1 (Winter 2006, Symposium: The Next Generation of Law School Rankings: Rankings Methodologies): 229–70.

library, and support services; student–faculty ratio; and the number of books in the law library) account for the last and smallest component.⁶⁵ As discussed in more detail by Espeland and Sauder, *USN* combines these various pieces of information into a single score, resulting in a ranking of all accredited schools. The *USN* law school rankings were first calculated and published in 1987, and have been published annually ever since.⁶⁶

Figure 2.2a captures the relationship between the stakeholder (sub)sets in 1987. As I noted in the previous section, the initial demand for this indicator came virtually entirely from magazine staff, especially senior editors seeking a way to boost circulation. Since the magazine also compiled the data on the various components, developed the algorithm for aggregating them, and literally supplied the indicator by publishing it, rule-demanders and rule-makers were at the outset virtually identical.

The intended users of the *USN* rankings were prospective law school students. While establishing the true intent of anyone with certainty may be impossible, especially ex post, the available information suggests that even the more optimistic among the *USN* senior staff expected only a small share of prospective law students to make use of the index, and they did not envision any other users. The target of the measurement (i.e., US law schools) was initially not even recognized as potential users, nor were law schools expected to react to the exercise through behavioral adaption—though they might have been recognized among the stakeholders had the editors thought about the issue in these terms. Notice that there is in this case virtually no overlap between the rule-makers and the users. Even though it might not have been fully recognized by the editors of *USN* at the time, this had the potential for a clearly asymmetric relationship of power.

Figure 2.2b captures the relationship in recent years.⁶⁷ The group of rule-demanders has grown tremendously, and now includes many users. While the users, especially prospective law students, still have little role in the rule-making, their demand now sustains the ranking to such an extent that the power relationship might be considered slightly less asymmetric (though the users face a far more

⁶⁵ The details are discussed in the “Law School Rankings Methodology” section of each ranking report, for 2011 written by Robert Morse and Sam Flanigan (available at <http://www.usnews.com/education/best-graduate-schools/articles/2011/03/14/law-school-rankings-methodology-2012_print.html>; last visited November 1, 2011). For more discussion see, e.g., Mitchell Berger, “Why the U.S. News and World Report Law School Rankings are Both Useful and Important,” *Journal of Legal Education* 51:4 (December 2001): 487–502, esp. 487f.

⁶⁶ Initially, only the top 25 schools were strictly ranked; the remainder was presented in tiers in an explicit recognition of the limits of precision—which appears to have disappeared over the years. See also Sauder and Lancaster, “Do Rankings Matter?” (n. 55); Wendy Nelson Espeland and Michael Sauder, “Rankings and Reactivity: How Public Measures Recreate Social Worlds,” *American Journal of Sociology* 113:1 (July 2007): 1–40; Michael Sauder and Wendy Nelson Espeland, “The Discipline of Rankings: Tight Coupling and Organizational Change,” *American Sociological Review* 74:1 (February 2009): 63–82.

⁶⁷ Changes in the size of each stakeholder subset circle are meant to indicate growth or decline in the size of this group of stakeholders over time, albeit not to scale: The outer circle (encompassing all stakeholders) is merely drawn to enclose the three or four subsets; any change in its size (or the amount of free space in that circle beyond the space occupied by the identified groups) is therefore not meant to be meaningful.

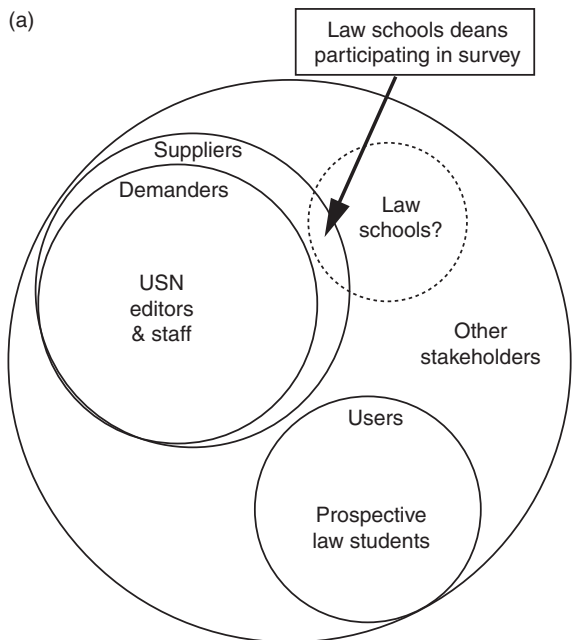


Fig. 2.2a *USN* Law School Rankings in 1987

Source: Tim Büthe, 2012.

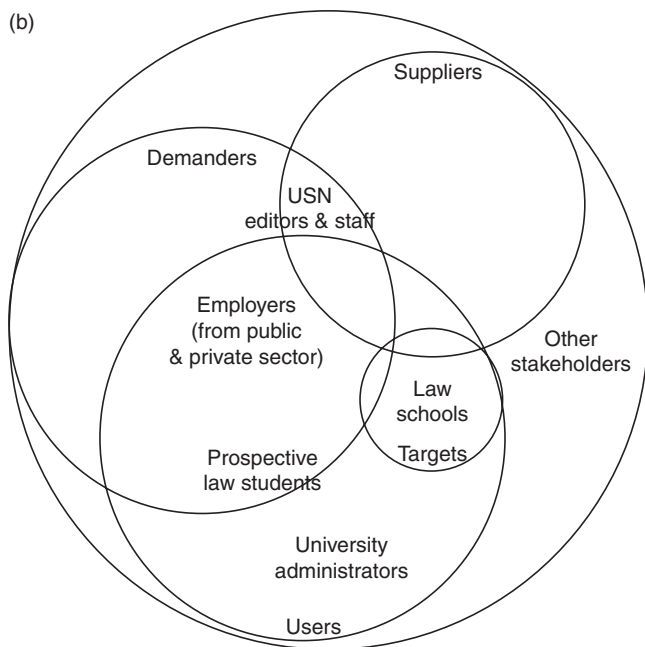


Fig. 2.2b *USN* Law School Rankings today

Source: Tim Büthe, 2012.

extreme collective action problem). The set of users is now much broader and now very prominently includes public- and private-sector employers who look to the rankings as an information shortcut for where the best students are likely to have gone.⁶⁸ Some of these “new” (or at least not initially intended) users are invited to participate in the rule-making through *USN*’s survey among legal practitioners. The group of users now also clearly includes the law schools themselves who, through the survey among the deans and selected faculty, play a small role in the rule-making but mostly find themselves in a very unequal relationship with the magazine.⁶⁹

Freedom in the World

The Freedom in the World ratings are the best-known output of the US-based NGO Freedom House (FH), which since 1941 has sought to support “the expansion of freedom in the world,” where freedom is not defined, but liberal democracy (and secondarily a market economy) is explicitly viewed as a prerequisite for it. As discussed in greater detail in the chapters by Dutta and Pistor,⁷⁰ the index has two major components: political rights and civil liberties. FH’s country experts assign to each country a score between 1 and 7, separately for each of the two components (lower numerical scores indicate better performance). A country’s overall score for political rights and for civil liberties, respectively, is supposed to measure these rights and liberties as actually experienced by individuals living under the political institutions that govern their societies, i.e., scores are supposed to reflect the actual operation of those institutions in practice, rather than just a country’s constitutional provisions, etc. Each score is based on a set of questions that ask FH’s coders to consider numerous different aspects of each of the two phenomena. To arrive at the political rights score, for example, FH asks coders to address nine questions, including: “Are the legislative representatives elected through free and fair elections?” and: “Is there a significant opposition vote . . . and a realistic possibility for the opposition to increase its support or gain power through elections?” Each coder is asked to assign points on each of the nine questions separately before assigning an overall political rights score for his/her country, but FH has steadfastly refused to make any disaggregated scores public.⁷¹

⁶⁸ Russell Korobkin, “In Praise of Law School Rankings: Solutions to Coordination and Collective Action Problems,” *Texas Law Review* 77:2 (December 1998): 403–28.

⁶⁹ Regarding the responsiveness of students—and in turn law schools—to the *USN* rankings, see also Espeland and Sauder, “Rankings and Reactivity” (n. 66); Sauder and Espeland, “The Discipline of Rankings” (n. 66).

⁷⁰ Nikhil K. Dutta, “Accountability in the Generation of Governance Indicators,” this volume, pp. 437–64 and Katharina Pistor, “Re-Construction of Private Indicators for Public Purposes,” this volume, pp. 165–79.

⁷¹ Each of the nine political rights and each of the 13 civil liberties questions has multiple sub-questions. The questions underpinning the civil liberties score include: “Are there free and independent media and other forms of cultural expression?,” “Are religious institutions and communities free to practice their faith and express themselves in public and private?,” and “Is there an independent judiciary?”

For each country, FH reports separate scores for political rights and political liberties, accompanied by a narrative report of the situation in each country. It then creates three groups of countries, based on the average of the two component scores (effectively collapsing a 13-category rating into a 3-category rating). The distinction between “free,” “partly free,” or “not free” countries (as shown in Table 2.1) is then reported as the overall result of the “Freedom in the World” survey. The rating has been widely interpreted as a rating of electoral + liberal democracy.

Who the original rule-demanders or even the original rule-makers for the Freedom in the World rating were is surprisingly unclear. Academic studies, including several careful comparative analyses of alternative measures of democracy, regularly note the lack of information about the origins and aggregation process of the Freedom House data,⁷² and there appear to be virtually no independent studies of the organization as such.⁷³ Freedom House provides minimal information about its funding in the early years of the Freedom in the World report, which was first published in 1973 (based on 1972 data), except to note that it received financial support from the US government, private foundations such as the MacArthur Foundation, and private individuals. In recent decades, US government funding

Table 2.1 Freedom House, Freedom in the World Ratings, 2009

Rating	“Free”		“Partly Free”		“Not Free”	
	1.0–1.5	2.0–2.5	3.0–3.5	4.0–5.0	5.5–6.0	6.0–6.5
Countries	Andorra	Antigua & Barbuda	Albania	Bhutan	Afghanistan	Belarus
	Australia	Argentina	Bangladesh	Burkina Faso	Algeria	Burma
	Austria	Benin	Bolivia	Burundi	Angola	Chad

	UK	Trinidad	Tanzania	Tonga	Vietnam	Syria
	USA	Ukraine	Turkey	Uganda	Yemen	Turkmenistan
	Uruguay	Vanuatu	Zambia	Venezuela	Zimbabwe	Uzbekistan

⁷² See, e.g., Kristian S. Gleditsch and Michael D. Ward, “Double Take: A Re-Examination of Democracy and Autocracy in Modern Politics,” *Journal of Conflict Resolution* 41:3 (June 1997): 361–83; Kenneth A. Bollen and Pamela Paxton, “Subjective Measures of Liberal Democracy,” *Comparative Political Studies* 33:2 (February 2000): 58–86; Zachary Elkins, “Gradations of Democracy? Empirical Tests of Alternative Conceptualizations,” *American Journal of Political Science* 44:2 (April 2000): 293–300; Gerardo L. Munck and Jay Verkuilen, “Conceptualizing and Measuring Democracy,” *Comparative Political Studies* 35:1 (February 2002): 5–34; Gretchen Casper and Claudiu Tufis, “Correlation Versus Interchangeability: The Limited Robustness of Empirical Finding on Democracy Using Highly Correlated Data Sets,” *Political Analysis* 11:2 (May 2003): 196–203.

⁷³ Diego Giannone (“Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case,” *Democratization* 17:1 (February 2010): 68–97) goes furthest when he grounds his analysis of the ideology of FH in a sketch of FH’s history, but even that history is very rudimentary and not based on primary sources.

has accounted for 80 percent or more of Freedom House's budget, mostly from the State Department, much of it via the National Endowment for Democracy and US AID. This has led many to portray the Freedom in the World rating as essentially responding to a demand from the US government,⁷⁴ which FH has vehemently denied (along with most other critiques) though it has provided little evidence to dispel the claim.⁷⁵

Notwithstanding such questions, the FH ratings have clearly been widely used and by all indications have been influential, including in public discourse and, according to Freedom House, also among multinational corporations assessing political risks. For recent years, many of the users of the index, especially journalists but also academic and policy researchers, might also be considered rule-demanders. Both the user and demander group thus clearly have grown over time. Figure 2.3 seeks to capture this configuration of stakeholders in recent years.⁷⁶

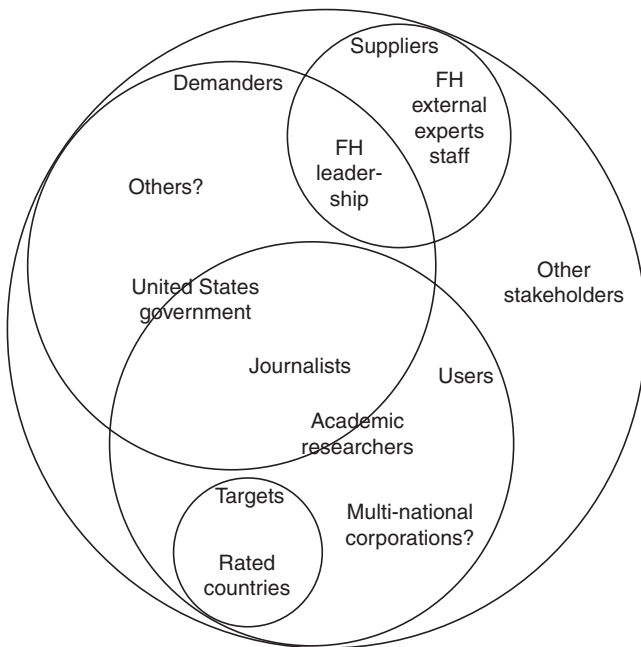


Fig. 2.3 FH Freedom in the World Ratings

Source: Tim Büthe, 2012.

⁷⁴ Initial uncertainty whether the index would have any impact on countries that were not assigned perfect scores could lead one to speculate whether it was in the earliest years an exercise in declamatory politics, designed to satisfy a demand for castigating countries with poor political rights or civil liberties (including many Cold War allies) without actually addressing the situation. This would imply that the user and target groups might be left out if Figure 2.3 were re-drawn for the earliest years.

⁷⁵ See, e.g., Raymond D. Gastil, "The Comparative Survey of Freedom: Experiences and Suggestions," *Studies in Comparative International Development* 25:1 (1990): 25–50.

⁷⁶ As Nikhil Dutta notes in his chapter, many unanswered questions surround Freedom House's claim that its country experts draw, inter alia, on academic studies as well as reports from NGOs from the country in question, and that this should be considered a way of allowing stakeholder participation.

Conclusion

This chapter has argued that the supply and use of indicators virtually always entails relationships of power. Models of a voluntary exchange relationship, which assume the full separation of suppliers and demanders,⁷⁷ as well as the ideal typical model of self-regulation, which assumes the two groups to be identical, inherently obscure these power relations, even though they are pervasive in global governance. To make these power relations more apparent, I have presented a conceptual framework that moves beyond models of supply and demand with a straightforward equilibrium between them and instead distinguishes four main groups of stakeholders. To illustrate the utility of the model, I have applied it to the examples of the *US News & World Report* ranking of US law schools and Freedom House's rating Freedom in the World. The conceptual model has various possible uses, including normative analysis. For positive analysis, it draws our attention to the challenging task at hand when we seek to understand the operation of an indicator: the greater the divergence of the stakeholder groups, the greater is the need to separately explain the incentives for demand, supply, and use of an indicator, as well as to explain the behavioral adaptation by the measured.⁷⁸ More broadly, the conceptual model helps us understand both the potential benefits and limitations of global governance through indicators.

Publishing long lists of "selected sources" that FH claims its experts have "consulted" amounts to little more than saying: "Trust us, we know these countries." I therefore do not show users as having any role in rule-making. The delegitimization of FH's measures through sustained critique, especially for failure to follow fundamental rules of scientific measurement (replicability in particular), might mean that the number of academic rule-demanders and maybe even the group of users is actually shrinking.

⁷⁷ Demanders here in the standard economics sense of purchaser-users. I have, by contrast, used "demanders"—in a more political sense—to denote all individuals and groups who either overtly call for an indicator or value it to the point where they are willing to give credit or pay some cost for its provision.

⁷⁸ I have also offered a number of possible explanations for each of them, without any claim to comprehensiveness.

3

Taming and Framing Indicators: A Legal Reconstruction of the OECD's Programme for International Student Assessment (PISA)*

Armin von Bogdandy and Matthias Goldmann

Indicators as a challenge for public law

The modern welfare state depends to a considerable extent on numerical data. Without statistical information, it is difficult to identify problems, draft policies in response, and measure progress. Ensuring the availability of sound statistical information was therefore a primary concern of the pioneers of the welfare state in the 19th century.¹ After the Second World War, international institutions quickly developed into hubs of statistical information on all sorts of issues. A lot of this activity focused on macroeconomic issues where states are highly interdependent. It gave rise to some series of publications by international institutions, which use a large number of indicators in order to assess trends in the economy. Long-standing examples are the semi-annual Economic Outlooks published by the Organization for Economic Cooperation and Development (OECD), or the World Economic Outlooks compiled by the International Monetary Fund (IMF).²

With the surge of global governance after the Cold War, the informational activities of international institutions reached a new level. They began collecting and disseminating information on more and more policy areas hitherto believed to

* Much of this chapter draws on Armin von Bogdandy and Matthias Goldmann, "The Exercise of Public Authority through National Policy Assessment: The OECD's PISA Policy as a Paradigm for a New International Standard Instrument," *International Organizations Law Review* 5 (2008 (2009)): 241. Apart from the cited sources, the information contained in this article was retrieved through a series of interviews with OECD and government representatives. We are grateful for their cooperation. We extend our thanks to the organizers and participants of the workshop "Indicators as a Technique of Global Governance" at the NYU School of Law in September 2010, and in particular to Judge Sabino Cassese for his valuable comments on our chapter.

¹ Alain Desrosières, *The Politics of Large Numbers. A History of Statistical Reasoning* (Cambridge, MA: Harvard University Press 1998), 147ff. (on France and the UK); and 178ff. (on Germany and the US). In Germany, statistical assessments even played a role in building a nation out of a formerly feudal, religiously and economically heterogeneous society (ibid., 179).

² Armin Schäfer, *Die neue Unverbindlichkeit. Wirtschaftspolitische Koordinierung in Europa* (Frankfurt a. M.: Campus 2005), 92ff., 112ff.

fall within the domestic jurisdiction of states. They also changed the ways in which the data is presented. In a world that had ceased to be divided by opposing alignments, it became more acceptable to rank countries according to their performance in a given policy sector. Thus, the United Nations Development Programme (UNDP) developed the Human Development Index (HDI) indicating the level of development,³ and the World Health Organization (WHO) published a ranking of national health systems.⁴ The OECD developed an even more ambitious program in the field of education. As it lacked comparative data about the performance of school policies, it designed an international comparative survey that measures the educational achievements of 15-year-old secondary school students. The outcomes of this survey, called the Program for International Student Assessment (PISA), are published every three years in detailed reports. However, it was neither the nuanced, statistical information in the PISA reports nor the insightful analyses contained in additional reports that attracted the most attention; rather, it was their clumsy summary, expressed in a simple country ranking. It is to some extent due to these rankings that PISA became influential as a policy: it not only changed educational policy in some important ways, but arguably led to some shifts of power from national and sub-national parliaments and executives to international networks of highly specialized bureaucrats.

While the next section elaborates on the history, causes, and effects of PISA and the ways in which it changes the exercise and distribution of power, the principal goal of this chapter is to develop a legal conceptualization of PISA, and more generally, a methodology for framing indicators and other information produced and disseminated by international institutions from the viewpoint of public law. The choice of this focus follows from our understanding of the function of public law. In accordance with standard theories of liberal democracy, we assume that public law is a crucial requirement for the legitimate exercise of public authority, whether on the domestic or international level. Public law spells out the conditions and limits for the exercise of power. This idea of public law also explains why we focus on *public* indicators and information. Privately produced indicators, such as law school rankings, might require entirely different standards, since their authors enjoy fundamental rights.⁵

On the basis of this concept of public law, we argue that policy assessments by international institutions like PISA are a form of public authority. Even though the PISA reports do not necessarily entail direct consequences, such as an increase or decrease in budget allotments, they have a more indirect, communicative impact on the way in which educational policy is made in the participating states: PISA determines to a considerable extent our perception of concepts, problems, and solutions in the field of school education. We call this form of public authority “governance by information.” PISA thus needs to be framed by public law—more

³ Kevin Davis, Benedict Kingsbury, and Sally Engle Merry, “Indicators as a Technology of Global Governance,” *Law and Society Review* 36:1 (2012): 71–104, esp. 95–9.

⁴ World Health Report 2000, available at <http://www.who.int/whr/2000/en/whr00_en.pdf>.

⁵ Cf. Wendy Espeland and Michael Sauder, “The Dynamism of Indicators,” in this volume, pp. 86–109; see also below.

precisely, by public international law—in order to ensure its legitimacy and address any conflicts and contestations to which this shift of power to the international level might give rise. This is not only a hypothetical concern. While a weakly accountable international expert bureaucracy largely drove the development of PISA, government involvement in the implementation of PISA raises doubts as to whether PISA really offers what it promises—namely, a “sugarcoat-free” assessment of educational systems.

Our legal conceptualization of PISA focuses on identifying, in an abstract, general manner, the particular kind of instrument that PISA exemplifies, and on consolidating and further developing the legal rules applicable to such kinds of instruments. We believe that this focus on instruments, and not on actors, is more appropriate for a global arena in which public authority is exercised by a multiplicity of actors and fora. A legal standardization of the instruments for the exercise of public authority allows the linkage of comparable forms of power with a legal framework that is appropriate for the specific legitimacy challenge implied by this authority. Standardized instruments would occupy an intermediate level between general, abstract legal principles and the concrete rules of international law that govern a particular form of public authority. As such, they would provide a conceptual tool not only for ensuring the legitimacy of PISA, but also for transferring the knowledge and experience gained with this instrument to other policy areas and international institutions. Given that the recent economic crisis requires Western states to improve their performance in many respects, we believe that a standard instrument for comparative, empirical policy assessments like PISA would be useful to policymakers, while being sensitive to the contingencies and risks implied in any empirical data survey.

Case study: Genesis and operation of PISA

The road to PISA: A peculiar win-win situation

The idea of an international, comparative survey of educational achievements has its roots in certain developments in the United States and France during the 1980s. Galvanized by a report about the rather poor average performance of US high school students in the sciences, the Reagan administration realized a need for improvement.⁶ In the Cold War logic of the time, lagging behind in education came close to a kind of unilateral disarmament, threatening the United States’ technological superiority over the Soviet Union. However, the US federal government lacks competence in educational policy.⁷ It therefore decided to play a “two-level-game”⁸ and approached the OECD in order to set up international programs

⁶ National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (1983).

⁷ Cf. the enumerative list in the US Constitution, Art. 1 Sec. 8.

⁸ Cf. Robert Putnam, “Diplomacy and Domestic Politics: The Logic of Two-Level Games,” *International Organization* 42 (1988): 427.

that would create pressure for domestic educational reform.⁹ On the other hand, the left-wing French government was interested in making its educational system fairer and providing better chances for children with a working-class background. It hoped that international comparative data would demonstrate the selectiveness of its existing educational system.¹⁰ Thus, Cold War strategic concerns and French egalitarian politics, as different as they may be, worked together to increase the demand on the OECD for comparative data on educational systems.

Initially, however, neither the OECD secretariat, nor the other member states of the Organization were very receptive to the idea. The OECD's Centre for Educational Research and Innovation (CERI)¹¹ pursued a more input-oriented approach to education. In Germany, for example, comparative educational assessments had been met with suspicion ever since 1971 when the newly established Federal Ministry of Education and Research had used an OECD review of Germany's educational system in order to blame the *Länder* for poor performance.¹² To advance its plans, the United States felt compelled to threaten to withdraw from the OECD's educational program entirely if the idea of comparative data were not to be further pursued. The Organization unwillingly yielded to the wishes of its largest member and most important contributor.¹³

However, comparing educational systems proved quite difficult, not only because of the major differences that exist between the school systems of the OECD member states—sometimes even within one state—but also because of the lack of experience with indicators. Which indicators, input- or output-oriented ones, would allow measurement and comparison of school education levels? In order to improve the understanding of indicators, the CERI Governing Board, a subsidiary body of the OECD Council,¹⁴ set up a research project called "Indicators of Education Systems" (INES) in 1988.¹⁵ It consists of several working groups, called

⁹ Stephan Leibfried and Kerstin Martens, "PISA—Internationalisierung von Bildungspolitik. Oder: Wie kommt die Landespolitik zur OECD?" *Leviathan* 36 (2008): 3, at 7–8; Kerstin Martens and Klaus Dieter Wolf, "Paradoxien der Neuen Staatsräson. Die Internationalisierung der Bildungspolitik in der EU und der OECD," *Zeitschrift für Internationale Beziehungen* 13 (2006): 145, 165–6.

¹⁰ Martens and Wolf, "Paradoxien der Neuen Staatsräson" (n. 9), 166.

¹¹ The current mandate of CERI is contained in the Resolution of the Council approved at its 1046th session, December 14, 2006, C/M(2006)20, item 265; and C(2006)173.

¹² Georg Knauss, "The Role of Inspectorates and Country Policy Reviews," in OECD (ed.), *Knowledge Bases for Education Policies* (Paris: OECD, 1996), 86, 88.

¹³ Martens and Wolf, "Paradoxien der Neuen Staatsräson" (n. 9), 166.

¹⁴ We use the term "subsidiary body" for bodies formally established by statutory organs or other subsidiary bodies of the organization. For a critical discussion of the terminology see Santiago Torres Bernárdez, "Subsidiary Organs," in R.-J. Dupuy (ed.), *Manuel sur les organisations internationales* (Dordrecht: Nijhoff, 1988), 100, 101ff.

¹⁵ INES was established as a joint project to be financed by both the OECD and the member states (CERI Governing Board, Summary Record of the 38th Session (held at Paris on May 30–31, 1988), CERI/CD/M(88)1, October 25, 1988, para. 13). Since 2000, the various INES Networks have been coordinated by the INES Strategic Management Group, which was jointly established by the CERI Governing Board and the OECD Education Committee in order to achieve a more strategic and transparent management of INES (Decision of the Education Committee DEELSA/ED/M(2000)2, para. 18; decision of the CERI Governing Board CERI/CD/M(2000)2 para. 42). The Strategic Management Group is composed of four members of the CERI Governing Board and four of the OECD Education Committee, along with two additional members from the Employment, Labour

"Networks," each of which is composed of government representatives with a high level of expertise.¹⁶ Among these, Network A, which was charged with the development of indicators on the outcomes of education, acquired particular prominence.¹⁷ It provided the indicators used in the OECD's first comparative publication on education, *Education at a Glance*, in 1992. This soon became the OECD's flagship annual publication on education. The indicators used in the first few editions of *Education at a Glance* showed a considerable degree of experimentalism. For each subsequent edition, indicators were added and dropped, and some of them refined.¹⁸ Most of the indicators related to the resources and processes of education, i.e., the "hard" statistical facts, such as the financial means spent on education, or the percentage of an age group receiving secondary education.

However, it was soon realized that in order to accurately evaluate educational systems, one would have to possess data on students' actual performance. The OECD lacked reliable and comparative data on student achievements in order to produce meaningful indicators.¹⁹ The only available data on outcomes were the Trends in International Mathematics and Science Study (TIMSS) survey on competencies in mathematics and science and the Reading Literacy survey produced by the International Association for the Evaluation of Educational Achievement (IEA), a public-private hybrid involving governments as well as researchers from a large number of states.²⁰ This data was only of limited use for the purposes of the OECD. For example, the age of the students tested varied from survey to survey, and the testing intervals were rather large. Those involved in the production of *Education at a Glance* were therefore skeptical about the value of the performance indicators used in the first editions.²¹

This insight gave rise to the desire within the OECD Secretariat,²² as well as among INES members,²³ to produce its own regular, periodic source of data. The way to PISA was still rather long, however. The first plans for an empirical data survey that the Secretariat presented to the member states at the 1995 INES General Assembly in Lahti got a mixed reception. Five states were in favor of the proposal and some others showed cautious support, while a large number of states considered the project as too expensive, of limited political use, or an unwarranted

and Social Affairs Committee. The INES Strategic Management Group is supervised by a Joint Session of the CERI Governing Board and Education Committee.

¹⁶ On the composition and mandates of the INES Networks in its early years see National Research Council (Commission on Behavioral and Social Sciences and Education; Board on International Comparative Studies in Education), *International Comparative Studies in Education: Descriptions of Selected Large-Scale Assessments and Case Studies* (Winter 1994–5), 29–41.

¹⁷ Norberto Bottani, "OECD International Education Indicators," *International Journal of Educational Research* 25 (1996): 279.

¹⁸ Bottani, "OECD International Education Indicators" (n. 17), 280–3.

¹⁹ Bottani, "OECD International Education Indicators" (n. 17), 283.

²⁰ For a closer analysis of the IEA see Matthias Goldmann, "The Accountability of Private vs Public Governance 'by Information': A Comparison of the Assessment Activities of the OECD and the IEA in the Field of Education," *Rivista trimestrale di diritto pubblico* 58 (2008): 41.

²¹ Bottani, "OECD International Education Indicators" (n. 17), 284.

²² Martens and Wolf, "Paradoxien der Neuen Staatsräson" (n. 9), 167.

²³ INES Network A Meeting Record, Plenary Meeting, June 21–23, 1994, 3.

interference into a sensitive policy area.²⁴ Nevertheless, the OECD Secretariat and the supportive states interpreted these reactions as a mandate to continue with its plans for an empirical assessment.²⁵ This marked the beginning of the “data strategy” of INES Network A, the blueprint of PISA, which was developed between 1995 and 1997.

The OECD Secretariat strongly influenced the entire process. The Secretariat articulated expectations towards the development of the data strategy by the Network, submitted drafts, which the Network discussed, modified, and approved step by step. In order to advance the development of the data strategy, the Secretariat recruited staff formerly employed by the IEA,²⁶ among them Andreas Schleicher, a German mathematician with experience in empirical surveys, who should be considered the spiritual father of PISA. Dissatisfaction among the IEA staff stemming from the lack of political support for the IEA’s activities may have facilitated recruitment for the OECD.²⁷ Perhaps the OECD promised better opportunities for the realization of their visions.

Already by the end of 1995, the Secretariat had secured the unanimous support of INES Network A for its data strategy.²⁸ The final strategy, together with an elaborate proposal for the design of its management, was submitted to the OECD Education Committee and the CERI Governing Board in spring 1997 (the Strategy).²⁹ The Strategy suggested a so-called literacy concept for PISA. Accordingly, the survey would assess students’ mastery of processes, understanding of concepts, and ability to apply previously acquired knowledge, rather than the mere knowledge of content prescribed by the school curriculum. This was deemed necessary in light of the cultural and curricular differences among the students to be tested. Further, the Strategy proposed to measure how students’ skills relate to important demographic, social, and economic factors. The Strategy did not, however, spell out the details of the survey. The definition of the indicators to be used was left to the implementation stage. The Strategy only set out in general terms that the project should include basic indicators on knowledge, skills, and competencies; contextual indicators on demographic, social, educational, and economic variables; and trend indicators that reveal changes over time throughout the three envisaged survey cycles.³⁰ Basic indicators were to be reported by grouping students into different proficiency levels. Although the Strategy recognized the difficulties involved in the definition of proficiency levels, it expressed the belief that this way of reporting would be more useful for policymakers.³¹ With a view to strengthening the political salience of the project, the Strategy proposed in-

²⁴ Martens and Wolf, “Paradoxien der Neuen Staatsräson” (n. 9), 167; Leibfried and Martens “PISA—Internationalisierung von Bildungspolitik” (n. 9), 10.

²⁵ Cf. INES Compendium, Contributions from the INES Networks and Working Groups, document submitted to the Fourth General Assembly of the OECD Education Indicators Programme, Tokyo, GA(2000)12, September 11–13, 2000, Introduction, para. 1.

²⁶ Martens and Wolf, “Paradoxien der Neuen Staatsräson” (n. 9), 167.

²⁷ INES Network A Meeting Record, Plenary Meeting, March 28–31, 1995, 4.

²⁸ *Ibid.*, 2.

²⁹ OECD, A Strategy for Student Achievement Outcomes, DEELSA/ED/CERI/CD(97)4, March 28, 1997.

³⁰ *Ibid.*, paras 13–22.

³¹ *Ibid.*, para. 17.

depth policy analyses that go beyond mere statistical reporting and study, such as the impact of differences in education systems on student performance, gender proficiency, etc.³²

The OECD adopted the Strategy in several steps. After endorsement by the Education Committee and CERI Governing Board, the OECD Secretariat exchanged notes with member states by which it sought “a political decision for participation in the project as well as a commitment to contribute financially to the project.”³³ With this move, the organization wanted to make sure the project would be implemented before taking a decision on the level of the OECD. Once a sufficient number of commitments had been reached, the OECD Council adopted PISA with a binding resolution pursuant to Article 5(a) of the OECD Convention.³⁴ The choice of a binding resolution should enable the OECD to set up a separate budget for PISA.³⁵ In its material part, the resolution endorses the Strategy and the procedures set out there.³⁶

PISA in operation

The 1997 decision envisaged three subsequent assessment cycles, one every three years. The first assessment cycle carried out in 2000 focused on reading literacy, the second cycle in 2003 on mathematics literacy, and the 2006 assessment cycle on science literacy. In the same year, the Organization agreed to continue PISA with further triennial cycles.³⁷ The 2009, 2012, and 2015, assessment cycles repeat the three focus areas. Although run by the OECD, the PISA study is not limited to OECD member states. While two OECD members did not take part in the first assessment cycle,³⁸ four non-OECD states have participated in PISA from the beginning.³⁹ Eleven additional non-member states followed their example and took part in the first assessment cycle a year later.⁴⁰ In the second assessment cycle in

³² OECD, *A Strategy for Student Achievement Outcomes*, DEELSA/ED/CERI/CD(97)4, March 28, 1997, paras 23–5.

³³ OECD, *A Strategy for Producing Student Achievement Indicators on a Regular Basis*, Summary of Decisions Taken, Meeting in Budapest, Hungary, May 7–8, 1997, DEELSA/ED/CERI/CD(97)7, August 19, 1997, para. 19.

³⁴ OECD, *Resolution of the Council*, September 26, 1997, C(97)176/FINAL. Although the instrument is entitled “resolution,” the first paragraph of the preamble refers to Art. 5(a) of the OECD Convention, according to which the OECD may adopt “decisions” which have binding force. OECD practice uses both terms synonymously.

³⁵ Information retrieved through interviews; cf. OECD, *Resolution of the Council*, September 26, 1997, C(97)176/FINAL, para. 4.

³⁶ See para. 3 of the preamble of the resolution as well as the annexed document on the operational role of the Pisa Governing Board, which states that the programme should be carried out as described in the Strategy.

³⁷ OECD Council, *Resolution Concerning the Mandate of the Programme for International Student Assessment*, December 14, 2006, C(2006)173, Annex VI. The duration of the mandate is five years, which is the standard period for OECD projects.

³⁸ Slovak Republic, Turkey.

³⁹ Brazil, Latvia, Liechtenstein, and the Russian Federation.

⁴⁰ Albania, Argentina, Bulgaria, Chile, Hong Kong-China, Indonesia, Israel, FYR Macedonia, Peru, Romania, Thailand.

2003, a total of 11 non-member states participated from the beginning of the cycle.⁴¹ Taken together, the three assessment cycles cover one-third of the world population, and the states involved produce almost nine-tenths of the world's GDP.⁴²

Implementation has been steered during each project cycle by the Board of Participating Countries (BPC), subsequently renamed the PISA Governing Board (PGB). It is a subsidiary body of the OECD Council,⁴³ composed of one national expert from each participating country,⁴⁴ with the OECD Secretariat serving as the Secretariat of the PGB. At the beginning of its operation in 1997, a high personal overlap existed between INES Network A delegates and the PGB.⁴⁵ This transfer of institutional and subject matter knowledge reduced transition costs and ensured a smooth start of the project.

The main tasks of the PGB comprise overseeing project implementation and taking all policy decisions. For example, the PGB determines the policy objectives for each cycle, selects the areas and subjects to be tested, sets the priorities for indicators, analysis and instrument development, and guides the preparation of the reports at the end of each assessment cycle.⁴⁶ In particular, at the beginning of the first assessment cycle, the PGB took several decisions of fundamental importance which have since determined the shape of the project, and which cannot readily be reversed. They include the determination of the level of reliability desired for the study⁴⁷ and the adoption of proficiency levels.⁴⁸ The PGB also defined what it would understand as literacy in reading, mathematics, and science—the three domains of the assessment⁴⁹—and settled the very controversial question of the targeted age group⁵⁰—a decision that required balancing quality with the practicability of the assessment.⁵¹ In addition, the PGB determined in a general fashion what type of questions should be used (e.g., multiple-choice and open questions), and how the assessment materials should look.⁵²

Decisions of the PGB are binding on the participating states. They derive their binding force from the underlying Council resolution, which in turn is based on Article 5(a) of the OECD Convention. They require consensus, or, if brought to a

⁴¹ Brazil, Hong Kong-China, Indonesia, Latvia, Liechtenstein, Macao-China, Russian Federation, Serbia, Thailand, Tunisia, Uruguay.

⁴² OECD, *Learning for Tomorrow's World. First Results from PISA 2003* (Paris: OECD 2004), 20.

⁴³ It finds its legal basis in the resolution setting up PISA, *see n. 34*.

⁴⁴ OECD Council, Operational Role of the Board of Participating Countries, September 10, 1997, C(97)176, appendix.

⁴⁵ NWA-BPC Transition Meeting Record, May 7–8, 1997, 1.

⁴⁶ DEELSA/ED/CERI/CD(97)4, March 28, 1997, para. 72.

⁴⁷ INES Network, Plenary Meeting of October 28–30, 1996, Meeting Record, 4.

⁴⁸ Tenth Meeting of the BPC, Summary Meeting Record, DEELSA/PISA/BPC/M(2001)1, May 18, 2001, 8.

⁴⁹ Third Meeting of the BPC, DEELSA/PISA/BPC(98)8, April 20–21, 1998, 5. Accordingly, reading literacy is defined as “the knowledge and skills needed to understand, use and reflect on written and electronic texts, in order to participate in society, to achieve one's goals and to develop one's knowledge and potential.”

⁵⁰ INES Network, Plenary Meeting of October 28–30, 1996, Meeting Record, 4.

⁵¹ Third Meeting of the BPC, DEELSA/PISA/BPC(98)8, April 20–21, 1998, 8–10.

⁵² *Ibid.*, 6; Fourth Meeting of the BPC, DEELSA/PISA/BPC(98)16, July 6–7, 1998, 3.

vote, a two-thirds majority of the members of the board.⁵³ In practice, informal techniques are used in order to build consensus. For example, the secretariat may submit proposals by email that are accepted as soon as no one objects, be it with or without debate. Majority votes seem to occur, but they are not tracked in the meeting records.⁵⁴ Only rules and regulations concerning the operation of the PGB, changes to the formula used for calculating financial contributions, and changes to the project design—i.e., the strategy prepared by INES Network A, are excluded from majority vote.⁵⁵

Each assessment cycle starts with the issuance of an international tender for the international contractors (the Consortium), who develop and implement the assessment on the technical level. The Australian Council for Educational Research (ACER) was chosen by the PGB as the main international contractor for each cycle.⁵⁶ Each state also chooses a national contractor for each cycle who carries out the assessment on the local level. The actual development of each assessment is a very sensitive process involving many actors including the Consortium, Functional Expert Groups composed of internationally recognized scholars, a Technical Advisory Group including the Consortium and independent experts, as well as the PGB, which supervises the entire process and takes the main decisions referenced above.⁵⁷

Once the PGB has devised the assessment cycle in this way, Expert Groups develop assessment frameworks for each of the test domains and submit them to the PGB for adoption. These assessment frameworks specify what competencies should be tested in each domain, and how they should be weighted. For example, in the first assessment cycle it was decided that reading literacy should be tested along three dimensions ((1) context; (2) format and text type; (3) reading tasks or aspects, i.e., retrieving, interpreting and evaluating information).⁵⁸ On the basis of these directions, the Consortium develops the questionnaires. These have to undergo

⁵³ OECD Council, Operational Role of the Board of Participating Countries, C(97)176, September 10, 1997, appendix, para. 8; now OECD Council, Resolution Concerning the Mandate of the Programme for International Student Assessment, December 14, 2006, C(2006)173, Annex VI, "Decision-Making."

⁵⁴ On this point, we retrieved contradictory answers in our interviews.

⁵⁵ OECD Council, Operational Role of the Board of Participating Countries, C(97)176, September 10, 1997, appendix, para. 11; slightly rephrased in now OECD Council, Resolution Concerning the Mandate of the Programme for International Student Assessment, December 14, 2006, C(2006)173, Annex VI, "Decision-Making."

⁵⁶ E.g., Third Meeting of the BPC, DEELSA/PISA/BPC(98)8, April 20–21, 1998, 2.

⁵⁷ Functional Expert Groups on reading literacy, mathematics, sciences, as well as problem-solving were established in order to link the policy objectives specified by the PGB with the necessary scientific expertise (DEELSA/ED/CERI/CD(97)4, March 28, 1997, para. 78ff.). Each member of the expert groups was appointed as a contact point for several countries in order to ensure adequate representation of national interests. The Strategy stipulates a procedure for the appointment of the members of the Functional Expert Groups. Accordingly, the experts were to be nominated by the prime international contractor and appointed by the OECD Secretariat after consultation with the PGB. In practice, this procedure was modified in that the OECD Secretariat and the PGB played the main roles in the appointment procedure; see Third Meeting of the BPC, DEELSA/PISA/BPC(98)8, April 20–21, 1998, 3.

⁵⁸ Fifth Meeting of the BPC, October 29–30, 1998, DEELSA/PISA/BPC(98)32, p. 4; see also OECD, *PISA 2006—Science Competencies for Tomorrow's World* (2007), 284–5.

tests for cultural and gender neutrality.⁵⁹ For the first assessment cycle, the PGB decided to test the questionnaires beforehand in a small field study in order to gain insights about the difficulty of the questions.

For the actual assessment, the questionnaires need to be answered by a representative sample of about 5,000 students from each participating state. Schools have some discretion in selecting the students to be tested. Once the data is collected, each country has the possibility to review and comment on its data and approve of its further use. This approval is a prerequisite for the further use of data from a country.⁶⁰ In the next step, the Consortium establishes comparative data sets. In a first phase, each country is identified by a code that only the country knows.⁶¹ Thereafter, the BPC decides which indicators to use for the international PISA reports, which reveal the results. The Consortium provides drafts of the PISA reports. The PGB reviews these drafts meticulously, sometimes requesting modifications, and eventually approves them by consensus. This phase still involves important choices regarding the design of indicators. For example, in the first cycle, after the data had been collected, it turned out that the proficiency levels envisaged for the presentation of the results were not meaningful. The lowest proficiency level in reading literacy included students who showed at least some basic skills as well as students who failed to demonstrate the lowest skills that PISA measures. This had to be adjusted.⁶² Another major problem arose because the data sample collected in the Netherlands suffered from a low response rate.⁶³ Apart from these technical difficulties, the PGB also used its political discretion in this phase and ordered the Secretariat to, *inter alia*, make the initial PISA report more forward-looking and emphasize the effects of education for adult life.⁶⁴

The dissemination of the results follows an elaborate strategy developed by the PGB.⁶⁵ The OECD Secretariat has the privilege of being the first to publish the results. However, it follows from the consensual nature of the PISA reports and the care invested in their drafting that the Secretariat is not entitled to evaluate the data in ways other than those authorized by the PGB. Participating states are free to produce national reports once the first international report has been released.⁶⁶ Further thematic reports, as well as a detailed technical report by the OECD, provide more in-depth analyses.⁶⁷

What have been the effects of PISA on educational policy in the participating states? Before embarking on an analysis of those effects, we consider it useful to

⁵⁹ E.g., Seventh Meeting of the BPC, October 4–5, 1999, DEELSA/PISA/BPC/M(99)1, 3.

⁶⁰ DEELSA/ED/CERI/CD(97)4, March 28, 1997, para. 116; Ninth Meeting of the BPC, October 26–27, 2000, DEELSA/PISA/BPC/M(2000)2REV1, 8.

⁶¹ Ninth Meeting of the BPC, October 26–27, 2000, DEELSA/PISA/BPC/M(2000)2REV1, 8.

⁶² 11th Meeting of the BPC, July 16–18, 2001, DEELSA/PISA/BPC/M(2001)2, 4.

⁶³ *Ibid.*, 5.

⁶⁴ *Ibid.*, 7.

⁶⁵ Sixth Meeting of the BPC, March 1–3, 1999, DEELSA/PISA/BPC/M(99)1, 3–4; Seventh Meeting of the BPC, October 4–5, 1999, DEELSA/PISA/BPC/M(99)1, 3.

⁶⁶ *Ibid.*

⁶⁷ For the 2006 assessment cycle cf. <http://www.oecd.org/document/29/0,3343,en_2649_35845621_37563421_1_1_1_1,00.html>.

introduce our concept of international public authority. This is the threshold by which we measure the effects of PISA. As it is inextricably linked to our idea of public law, we start with an overview of the latter. Both explanations will be kept short as we have laid down our ideas elsewhere.⁶⁸

PISA as a challenge for public law?

The function of public law

Public law, at least in liberal democracies, is directly linked to the enlightenment idea of liberty. The general function of public law is to square liberty with the need of any society for the exercise of public authority, i.e., the making of unilateral decisions taken in the name and interest of an overarching, general entity, and their enforcement. The legal framework for decision-making processes resulting in such unilateral action needs to ensure that they are considerate of liberty.⁶⁹ The public law of liberal democracies does so by guaranteeing appropriate institutional settings, fair procedures, substantive standards, judicial review, and other forms of redress.⁷⁰ In other words, public law is about providing legitimacy to public authority by ensuring that it respects liberty. Of course, all these safeguards do not guarantee that the outcome will be just in a material sense or not yield to special interests. Even if all legal requirements are met, public authority will favor some interests over others. It is for exactly this reason that public law provides accountability and procedures to ensure that the previous decision can be reversed. Although a public law framework of authority does not automatically create legitimacy, it is a necessary precondition for legitimacy. In a liberal state, it is difficult to imagine legitimate public authority outside the structures of public law.⁷¹ In addition, the legal framework of public authority allows legitimacy not only to exist, but also to be seen and to be understood to exist.

The legitimacy-inducing function of public law is not only a necessity for the domestic level or acts of the European Union, but also for the public authority exercised by international institutions. Traditionally, such authority is justified by

⁶⁸ See, in particular, Armin von Bogdandy, Philipp Dann, and Matthias Goldmann, "Developing the Publicness of International Public Law: Towards a Legal Framework for Global Governance Activities," *German Law Journal* 9 (2008): 1375–400, as well as the contributions in Armin von Bogdandy et al. (eds), *The Exercise of Public Authority by International Institutions. Advancing International Institutional Law* (Heidelberg: Springer, 2010).

⁶⁹ E.g., Immanuel Kant, "Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht," in *Schriften zur Anthropologie, Geschichtsphilosophie, Politik und Pädagogik*, vol. 1 (Frankfurt a. M.: Suhrkamp, 1968 (orig. 1784)), 33, 39–41.

⁷⁰ Sabino Cassese, "Lo spazio giuridico globale," *Rivista trimestrale di diritto pubblico* 52 (2002): 323.

⁷¹ This idea of public law by and large follows the ideas of Hans Kelsen, *Allgemeine Staatslehre* (2nd edn, Berlin: Springer, 1925), 320–71. This is not to be confused with Weber's idea of legitimacy through legality. Weber assumed that the mere existence of rational-bureaucratic decision-making framework evokes legitimacy, irrespective of its quality; see his *Wirtschaft und Gesellschaft*, vol 1 (5th edn, Tübingen: Mohr Siebeck, 1972 (orig. 1925)), § 7.

the principle of state consent. Ensuring this principle requires legal structures. However, more and more doubts are being raised about the capability of this principle to render international public authority legitimate. International public authority often disturbs domestic balances of power and dilutes domestic accountability mechanisms set up to respond to the idea of liberty. For example, decisions might effectively be taken by global epistemic communities, which are hard to control.⁷² Not to participate in global regulatory efforts is no option either as it entails serious economic and political risks for a state. These growing concerns about unaccountable international bureaucracies gave rise to a debate on an emerging worldwide constitution or Global Administrative Law.⁷³ At the same time, these concerns are a sign of the success of international institutions as efficient solvers of global problems that each state alone could not tackle. In this situation, we think the most promising strategy is to enable the legitimate exercise of such international public authority by advancing international institutional law and elaborating sufficiently complex conceptual frameworks for the exercise of public authority.

Methodologically, the development of such a conceptual framework is a dialectical process that we call doctrinal constructivism. In the Western legal tradition, with the exception of Britain, legal scholarship plays a central role in this process.⁷⁴ Scholarly analysis links the practice of international institutions with deductive reasoning guided by legal principles and considerations of legitimacy. The resulting conceptual suggestions may be adopted by legal practice. While courts are the principal implementers of scholarly concepts on the national level, doctrinal constructivism on the international level relies on an array of actors, including international institutions, governments, parliaments, and international and domestic courts. Such practical applications are in turn subject to scholarly criticism. This multi-player, communicative ping-pong game adjusts and stabilizes the conceptual proposals. What used to be a mere suggestion, a communicative reference point, is slowly considered part of the existing law.⁷⁵ In the optimal case, the result leads to practically reasonable solutions and enjoys wide acceptance. The concept of treaty in international law is such a success story,⁷⁶ just as that of the administrative act in continental public law.⁷⁷

⁷² Peter M. Haas, "Epistemic Communities and International Policy Coordination," *International Organization* 46 (1992): 1.

⁷³ E.g., Benedict Kingsbury, Nico Krisch, and Richard Stewart, "The Emergence of Global Administrative Law," *Law and Contemporary Problems* 68 (2005): 15; Anne Peters, "Compensatory Constitutionalism: The Function and Potential of Fundamental International Norms and Structures," *Leiden Journal of International Law* 19 (2006): 579.

⁷⁴ On the legacy and significance of doctrinal constructivism in a comparative perspective, see Armin von Bogdandy, "The Past and Promise of Doctrinal Constructivism: A Strategy for Responding to the Challenges Facing Constitutional Scholarship in Europe," *International Journal of Constitutional Law* 7 (2009): 364, esp. 376ff.

⁷⁵ Wolfgang Hoffmann-Riem, "Methoden einer anwendungsorientierten Verwaltungsrechtswissenschaft," in Eberhard Schmidt-Aßmann and Wolfgang Hoffmann-Riem (eds), *Methoden der Verwaltungsrechtswissenschaft* (Baden-Baden: Nomos, 2004), 9, 18.

⁷⁶ Milos Vec, *Recht und Normierung in der Industriellen Revolution* (Frankfurt a.M.: Klostermann, 2006), 105–20.

⁷⁷ von Bogdandy and Goldmann, "The Exercise of Public Authority through National Policy Assessment" (n. 68), 271–81.

In developing conceptual proposals, legal scholarship draws legitimacy from reference to authoritative documents and practice, as well as from rational arguments that meet the established standards of the discipline.⁷⁸ In international law, this role of scholarship is recognized in Article 38 para. 1 lit. d of the Statute of the International Court of Justice. Nonetheless, legal scholarship is far from an objective craft, free of the contingencies implied in any argument. An enlightened form of doctrinal constructivism needs to make these contingencies as visible and transparent as possible and spell out their normative foundations.⁷⁹ As said above, our normative starting point is the respect for liberty, and, as a consequence of that, for individual rights. If one assumes that constitutional principles are emerging in international law, respect for liberty and individual rights will probably be among them.⁸⁰ It should be emphasized that liberty and individual rights do not necessitate maximum limitations on the power of international institutions. Rather, they require effective international institutions wherever state performance is low or where a single state cannot provide important public goods. Conversely, the capacity of states and the legitimacy of their structures can be useful prerequisites for the realization of liberty and individual rights. Figuring out an acceptable balance between liberty and efficiency, or between the state and international institutions, requires an interdisciplinary approach that takes a closer look at the social context and effects as well as at the legitimacy of the forms of public authority in question. This contrasts our enlightened version of doctrinal constructivism with the disciplinary isolationism of nineteenth-century “Begriffsjurisprudenz.”

The concept of international public authority

Developing a legal framework for public authority requires as a first logical step the identification of acts that infringe liberty or individual rights. At present, there is no generally recognized definition of international public authority. In many disciplines, law and social sciences alike, authority has long been associated with the capacity to issue orders backed by physical sanctions.⁸¹ Under these terms, PISA could certainly not be qualified as public authority. Although the PISA reports contain information that might have normative *effects*, they do not include explicit or implicit orders addressed to the participating states.

⁷⁸ On the legitimating force of intersubjectively acceptable rational arguments in the context of legal interpretation, Ralph Christensen and Hans Kudlich, *Gesetzesbindung: Vom vertikalen zum horizontalen Verständnis* (Berlin: Duncker & Humblot, 2008), 195–218, esp. 198–202.

⁷⁹ Similar Andreas Paulus, “International Law After Postmodernism: Towards Renewal or Decline of International Law?” *Leiden Journal of International Law* 14 (2001): 727.

⁸⁰ Cf. Stephan Kadelbach and Thomas Kleinlein, “International Law—a Constitution for Mankind? An Attempt at a Re-appraisal with an Analysis of Constitutional Principles,” *German Yearbook of International Law* 50 (2007): 303, 319–29.

⁸¹ John Austin, *The Province of Jurisprudence Determined* (Cambridge: Cambridge University Press, 2001 (orig. 1832)), 21; Hans Kelsen, *Reine Rechtslehre* (2nd edn, Vienna: Deuticke, 1960), 34ff.; Max Weber, *Wirtschaft und Gesellschaft* (n. 71), 122; Robert Dahl, “The Concept of Power,” *Behavioral Science* 2 (1957): 201, 202; Ralph Dahrendorf, *Über den Ursprung der Ungleichheit unter den Menschen* (Tübingen: Mohr Siebeck, 1966), 20.

However, this narrow definition of public authority rests on problematic anthropological foundations. It is steeped in the early modern Christian idea of the corrupt nature of man (*natura corrupta*), which is believed to result from the fall of mankind.⁸² With Hobbes' "*homo homini lupus*," the idea was secularized and became one of political philosophy.⁸³ Accordingly, the threat of physical sanctions is an indispensable prerequisite for inducing compliance with a legal rule. In light of more modern empirical and theoretical research, we have good reason to believe that this is an overly restrictive perspective on the factors that determine human behavior. There is a large host of epistemological approaches demonstrating the diversity of factors that may have an impact on how states behave. For example, in a rational perspective, a rule or a piece of information might build up communicative pressure which the addressee can only avoid at some cost, be it reputational, economic, or other.⁸⁴ Constructivism points out the significance of communication for identity-formation and thus for decision-making,⁸⁵ and behavioral economics stresses the importance of cognitive frames in receiving information and taking action.⁸⁶ In light of this, it would be difficult to argue that information is neutral and separate from power relationships. Michel Foucault's examination of prison systems provides evidence to the contrary.⁸⁷ More recently, Alain

⁸² "Sed ubi pecasset, corrupta est voluntas et omnia Naturalia, et credo, quod corrupti etiam sint sensus naturales et corpus vitium in ipsis sensibus, sanguine, nervis, vidit Adam ante lapsum fortius." [Since the first commission of a sin, the will and entire physical nature of man have been corrupted, and I believe that his natural senses and the innocence of his body reflected in his senses, blood and nerves are also corrupt, Adam seeing stronger before the fall.], Martin Luther, *Edited Works*, part 1, vol 40, partition 2 (Weimar: Bohlaus, 1914), 323,10–324,3. This contrasts with the medieval idea of the double nature of man, cf. Thomas Aquinas, *Summa theologiae*, vol. 2 (Rome: Forzani, 1894 (orig. 1265–1273)), I-II, q. 109, Art. II., 829: "Respondeo dicendum, quod natura hominis potest *dupliciter* considerari: *uno modo* in sui integritate, sicut fuit in primo parente ante peccatum: *alio modo*, secundum quod est corrupta in nobis post peccatum primi parentis." [I say in reply that human nature can be considered in two ways: first in its integrity, as it became apparent in the beginning before the fall; second in respect of its corrupt element, which first appeared after the fall.]

⁸³ T. Hobbes, *De Cive* (Oxford: Clarendon Press, 1983 (orig. 1647)), Epistola dedicatoria, § [1].

⁸⁴ See Andrew Guzman, *How International Law Works. A Rational Choice Theory* (Oxford: Oxford University Press 2008); Dinah Shelton (ed.), *Commitment and Compliance. The Role of Non-binding Norms in the International Legal System* (Oxford: Oxford University Press, 2000); in the context of German administrative law M. Fehling, "Informelles Verwaltungshandeln," in W. Hoffmann-Riem et al. (eds), *Grundlagen des Verwaltungsrechts*, vol. 2 (Munich: Beck, 2008), 1341–404, at marginal note 7.

⁸⁵ Michael Barnett and Raymond Duvall, "Power in Global Governance," in M. Barnett and R. Duvall (eds), *Power in Global Governance* (Cambridge: Cambridge University Press, 2005), 1–32. Barnett and Duvall define power in very broad terms as "the production, in and through social relations, of effects that shape the capacities of actors to determine their own circumstances and fate" (*ibid.*, 8). Their research on power applies *mutatis mutandi* to our concept of authority. On the power implied in identity formation see Tony Porter and Michael Webb, "The Role of the OECD in the Orchestration of Global Knowledge Networks," paper presented at the Canadian Political Science Association Annual Meeting (2007), available at <<http://www.cpsa-acsp.ca/papers-2007/Porter-Webb.pdf>>.

⁸⁶ For example, whether an event is causing a loss or a benefit is not an ontological question, but depends on the cognitive attitude of those who assess it. See Daniel Kahneman and Amos Tversky, "Choices, Values, and Frames," in Daniel Kahneman and Amos Tversky (eds), *Choices, Values, and Frames* (Cambridge: Cambridge University Press, 2000), 3–16, 4.

⁸⁷ Michel Foucault, *Discipline and Punish. The Birth of the Prison* (New York: Random House, 1977), 27–8.

Desrosières' brilliant study of the rise of statistical information in nineteenth-century welfare states revealed that the compilation of the first census data entailed a new definition of many aspects of society, thus shaping further action and even the identity of nation states.⁸⁸ Consequently, the question whether an instrument of an international institution is authoritative cannot be decided in a black-and-white fashion according to whether it entails binding legal commands. Rather, there is a large grey area of instruments below this threshold that influence those to whom they are addressed or those who use them to a greater or lesser degree. It is the challenge for the concept of international public authority to identify authoritative instruments within this grey area.

Moreover, the concept needs to be attentive to the particularities of the international level. Simple domestic analogies would, for example, overlook the fact that actors on the international level are still very different from those on the national level, even though states are by far not the only international actors. For example, some policy areas on the international level are in the hands of a small group of highly specialized professionals representing a few stakeholders. Such a group might be more susceptible to soft, informal modes of governance than a large, amorphous and anonymous group of individuals like the addressees of domestic traffic regulation.

As a result, we define *authority* very broadly as the capacity based in law to *determine* the behavior of individuals, to reduce their liberty. First, such a determination can be a *legal* one. This is the case if an act modifies the rights and duties of a legal subject without its immediate consent. Legal acts, however, do not need to be backed by enforcement powers. Rather, we see the main function of law as the institutional creation and stabilization of normative expectations.⁸⁹ Second, a determination can also occur through an act that merely *conditions* other legal subjects. In a rational perspective, this is the case whenever that act builds up communicative pressure exceeding the threshold above which a legal subject concerned by the act cannot ignore it without serious consequences. In a non-rational perspective, this happens if the act carves out the cognitive environment of the issue concerned in a manner that marginalizes alternative perspectives.

As the public and international character of PISA do not raise grave problems, it suffices to reiterate only the basic traits of these two components of our definition of international public authority:⁹⁰ we consider as public and international any authority which is exercised on the basis of a formal or informal international act of public authorities, mostly states, to further a goal which they define, and are authorized to define, as being of public interest. The public and international character of an authoritative act is first and foremost determined by its legal basis. This allows for greater clarity and also for distinguishing public from private acts.

⁸⁸ Desrosières, *The Politics of Large Numbers* (n. 1), 179. This applies in particular to Germany.

⁸⁹ This is probably the smallest common denominator of otherwise incompatible theories such as those by Jürgen Habermas and Niklas Luhmann, respectively. See Jürgen Habermas, *Faktizität und Geltung* (Frankfurt a. M.: Suhrkamp, 1992), 90–108, esp. 106, 146; Niklas Luhmann, *Das Recht der Gesellschaft* (Frankfurt a. M.: Suhrkamp, 1993), 150–3.

⁹⁰ For a more detailed discussion see von Bogdandy, Dann, and Goldmann, "Developing the Publicness of International Public Law" (n. 68), 1383–5.

Private acts derive legitimacy from individual liberty, whereas public authority enjoys no similar a priori legitimacy.⁹¹

The impact of PISA: Governance by information as public authority?

PISA as “governance by information”

Applying the above definition, this section asks whether PISA can be considered as having a relevant impact on liberty in order to be qualified as international public authority. In order to answer this question, it is important to get a better understanding of the type of governance that has been used for PISA, which we call “governance by information.” Governance by information determines a given policy field by creating pressure on, or shaping the cognitive framework of, policy-makers and policymaking through the collection, processing, and dissemination of information.⁹² It is thus not immediately focused on legal standards, like reporting mechanisms of international treaties which are focused on the question whether a state complied with a pre-established standard. Rather, governance by information may implicitly or explicitly generate new standards or expand existing standards. The distinctive element of governance by information is that the information it produces is not coded according to the binary distinction between legal and illegal. Also, governance by information typically prefers long-term impacts over short-term effects.

International and supranational institutions more and more often have recourse to governance by information. Other than PISA, examples include the OECD Economic Outlooks, the Open Method of Coordination of the European Union,⁹³ or the Assessment Reports by the Intergovernmental Panel on Climate Change.⁹⁴ However, even treaty review mechanisms may today bear traits of governance by information. For example, the Committee for Economic, Social and Cultural Rights uses indicators in order to assess states’ progress in realizing these rights.

⁹¹ Christoph Möllers, *Die drei Gewalten. Legitimation der Gewaltengliederung in Verfassungsstaat, Europäischer Integration und Internationalisierung* (Weilerswist: Velbrück, 2008), 71–3.

⁹² The terminology and definitions used for this and similar governance modes varies. “Governance by persuasion” is used by Niklas Noaksson and Kerstin Jacobsson, “The Production of Ideas and Expert Knowledge in OECD. The OECD Jobs Strategy in Contrast with the EU Employment Strategy,” SCORE rapport series No. 2003/7, Stockholm, 32–4; cf. however Dirk Lehmkuhl, “Governance by Rating and Ranking,” paper presented at the International Studies Association Annual Meeting (2005), on file with the authors. Martens et al. use the term “governance by opinion formation”; see Kerstin Martens, Carolin Balzer, Reinhold Sackmann, and Ansgar Weymann, “Comparing Governance of International Organisations—The EU, the OECD and Educational Policy,” Transtate Working Papers No. 7 (2004), available at <<http://www.sfb597.uni-bremen.de/homepages/martens/download.php?ID=8&SPRACHE=DE&TABLE=AP&TYPE=PDF>>. Elsewhere, Martens speaks of “governance by comparison,” cf. Kerstin Martens, “How to Become an Influential Actor—The ‘Comparative Turn’ in OECD Education Policy,” in Kerstin Martens et al. (eds), *New Arenas of Education Governance* (Basingstoke: Palgrave MacMillan, 2007), 40–56, 41.

⁹³ Presidency Conclusions, Lisbon European Council, March 23–24, 2000, available at <http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/cc/00100-r1.en0.htm>.

⁹⁴ <<http://www.ipcc.ch>>.

Indicators enable a surveillance mechanism that is less controversial than mechanisms operating on the basis of the legal/illegal distinction.⁹⁵

Given this understanding of the philosophy behind governance by information, an assessment of the impact of PISA on policymaking should focus more on long-term developments, such as changing attitudes and educational reforms, and less on short-term effects, such as immediate reactions to the PISA reports or their use in the attribution of funds. We do not think that press coverage of the release of a new PISA report is a reliable indicator for the impact of PISA.⁹⁶ Press coverage might depend on whether the PISA results exceed or disappoint public expectations about student performance.⁹⁷ PISA might also stir less interest in a state which has already put in place a sophisticated testing system of its own so that people are used to educational assessments.⁹⁸ Likewise, even though there is more than only anecdotal evidence that some governments specifically train their students' ability to handle PISA questionnaires,⁹⁹ such immediate consequences are not what governance by information is about.

PISA as international public authority?

In our opinion, a more reliable indicator for the impact of PISA, and thus for answering the question whether it is an act of international public authority, is to look at government reform projects that can be identified as direct or indirect consequences of PISA, and to examine whether and how they differ from past educational policies. To illustrate such a qualitative approach in limited space, we focus on reform projects in Germany only.

Germany is a case in point: the mediocre results of the first as well as the subsequent PISA reports triggered a public outcry that has become known as the "PISA shock." After some first, hectic reactions, this opened the door for substantial reform projects, all of which make implicit or explicit reference to PISA. The federal government set up an educational programme with a budget of no less than

⁹⁵ AnnJanette Rosga and Margaret Satterthwaite, "The Trust in Indicators: Measuring Human Rights," *Berkeley Journal of International Law* 27 (2009): 253; Michael Duchstein, *Das internationale Benchmarkingverfahren und seine Bedeutung für den gewerblichen Rechtsschutz* (Heidelberg: Springer, 2010).

⁹⁶ For such an approach, see Kerstin Martens and Dennis Niemann, "Governance by Comparison—How Ratings & Rankings Can Impact National Policy Making in Education," paper presented at the International Studies Association Annual Meeting, New York (2009), on file with the authors.

⁹⁷ The media reaction to PISA was very strong in Germany, where the PISA results came as a great disappointment (below). However, reactions were also strong in the PISA winner state Finland; see Sotiria Grek, "Governing by Numbers: The PISA 'Effect' in Europe," *Journal of Education Policy* 24 (2009): 23, 28–9. For a comparative analysis, see OECD INES Network A, Review of Assessment Activities, issue 16, February/March 2004, 2–4.

⁹⁸ At the example of the United Kingdom: Grek, "Governing by Numbers" (n. 97), 34.

⁹⁹ E.g., Hessisches Kultusministerium, "Materialsammlung und Übungsaufgaben des Instituts für Qualitätsentwicklung," press briefing, August 17, 2006, available at <<http://www.kultusministerium.hessen.de>>.

4 billion euros.¹⁰⁰ The *Länder* initiated substantive, long-term reform projects as well.¹⁰¹ Two distinguishable trends among these reform projects show a change in education policy resulting from PISA.

First, some reform projects aim at developing a more contextual way of learning and teaching, such as “Physics in Context” or “Chemistry in Context.”¹⁰² They bring together teachers and specialists to elaborate new teaching methods that enable students to better master their knowledge, to apply it to new contexts, and to expand it autonomously. These teaching methods are intended to replace previous, more confrontational and reproductive teaching styles. This is a direct consequence of PISA, which revealed the need to improve teaching methods in this direction; the projects also follow a concept of education that correlates with that used for PISA, which tests students’ ability to apply previously acquired knowledge.¹⁰³

Secondly, and more importantly, PISA succeeded in shifting the general approach to school education in Germany from one grounded in an input-oriented, normative, and philosophical understanding, to one based on an output-oriented, empirical, and economic understanding.¹⁰⁴ Previously, German schools cultivated a tradition of the “closed classroom door.”¹⁰⁵ Teachers based their professional identity on their autonomy, which included the freedom from accountability regarding the outcomes of education. They called this “pedagogic liberty,” a term which not by coincidence appeals to the language of fundamental rights, even though this does not reflect the legal situation accurately. As a result, school administrations confined themselves to outlining a broad framework for the inputs of school education. They left the implementation of this framework entirely to the talent of each teacher. This approach goes back to the educational ideal of Wilhelm von Humboldt, who saw the purpose of education in the formation of a self-determined individual and not in economic ends. Prescribing the outcomes of education would run diametrically against this idea as it sees education as an open process. This ideology was cemented by the fact that the administrators of the

¹⁰⁰ See “Verwaltungsvereinbarung Investitionsprogramm “Zukunft Bildung und Betreuung,” April 29, 2003, available at <http://www.bmbf.de/pub/20030512_verwaltungsvereinbarung_zukunft_bildung_und_betreuung.pdf>.

¹⁰¹ E.g., “Rahmenplanung für die Qualitätsentwicklung der Schulen im Land Bremen” (2006), available at <<http://www.bildung.bremen.de/fastmedia/13/rahmenplanung.pdf>>, which contains numerous references to PISA; on the reform projects of the Land Baden-Württemberg see “Reformprojekte in Baden-Württemberg,” undated document, available at <http://www.km-bw.de/servlet/PB/-/nbi9mb1bqwm151wlhin04968iu1bjg1lf/show/1101334/PISA-E_Reformvorhaben_Internet.pdf>.

¹⁰² On these and other projects, see Grek, “Governing by Numbers” (n. 97), 29; Hubert Ertl, “Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany,” *Oxford Review of Education* 32 (2006): 619, 623.

¹⁰³ Jürgen Baumert et al., “Was messen internationale Schulleistungsstudien?—Resultate kumulativer Wissenserwerbsprozesse,” *Psychologische Rundschau* 58 (2007): 118.

¹⁰⁴ Martens et al., “Comparing Governance of International Organisations” (n. 92).

¹⁰⁵ Hermann Lange, “PISA und kein Ende: Was bedeuten die Untersuchungen für die Schulverwaltung?,” *Recht der Jugend und des Bildungswesens* 51 (2003): 193.

educational system for the largest part had made their careers within the system and shared the ideology.¹⁰⁶

PISA and the ensuing public outcry in Germany led to a fundamental change in this attitude. Soon after the publication of the first PISA report, the federal and state governments took measures aimed at a reorientation of Germany's educational system towards educational outputs. A key element of this reform consists in the introduction of educational standards. Unlike the former input-oriented frameworks for learning, educational standards define the competencies which students are expected to master at a certain point in time. They imply an output-orientation as they are meant to serve both as a guideline for teachers and as a benchmark for the evaluation of educational achievements.¹⁰⁷ Educational standards were introduced at a speed that surprised observers, who had expected lengthy negotiations among the 16 state governments.¹⁰⁸ A 2003 feasibility study commissioned by the federal ministry for education and research provided the kick-off. Not surprisingly, it describes educational standards as an appropriate response to Germany's bad performance in PISA.¹⁰⁹ Since then, more and more educational standards have been adopted for different student groups and school types.¹¹⁰ With this rather rapid change from input to output, the anthropological foundations of Germany's educational system changed, too. Education ceased to be an end in itself and became a means for the accumulation of human capital. A utilitarian attitude that arguably neglects the humanities replaced Humboldt's approach to education.¹¹¹

After this rather German-centric analysis, it needs to be emphasized that PISA has caused major reverberations in other states, too,¹¹² although it triggered less public debate there. Even for high-performing states like Finland, PISA has been as much of a burden as an honor, since slipping from first rank in subsequent studies would be perceived as a defeat.¹¹³

In conclusion, taking into account the purpose of governance by information, which is about confirming or changing attitudes, PISA should be considered a form of international public authority. The cumulative effects of the appeal of an external, objective evaluation, and of "light" enforcement mechanisms are not unlikely to allow an instrument like PISA to have a significant influence on politics. It therefore requires a public law framework commensurate with the legitimacy challenges it involves. This raises the question, "How should a public law framework for PISA look?"

¹⁰⁶ Hermann Lange, "PISA und kein Ende: Was bedeuten die Untersuchungen für die Schulverwaltung?," *Recht der Jugend und des Bildungswesens* 51 (2003): 193.

¹⁰⁷ Ekkehard Klieme et al. (eds), *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung* (Berlin: Federal Ministry for Education and Research, 2003), 10 and 99.

¹⁰⁸ Ertl, "Educational Standards and the Changing Discourse on Education" (n. 101), 622.

¹⁰⁹ Klieme et al., *Zur Entwicklung nationaler Bildungsstandards* (n. 107), 11.

¹¹⁰ For a current list of agreed educational standards see <<http://www.bmbf.de/de/1154.php>>.

¹¹¹ On the neo-liberal concept of education within the OECD see Myriam Henry et al., *The OECD, Globalisation and Education Policy* (Amsterdam: Elsevier, 2001), 61ff., 102–5, 175.

¹¹² For a comparison see Grek, "Governing by Numbers" (n. 97).

¹¹³ Risto Rinne, Johanna Kallo, and Sanno Hokka, "Too Eager to Comply? OECD Education Policies and the Finnish Response," *European Educational Research Journal* 3 (2004), 454.

Developing a legal framework for PISA

In spite of the enormous impact of PISA on national policy, there has been little exploration of how PISA should be qualified in legal terms and what legal standards should be applied to it. We hypothesize that this is because the knowledge and experience of international lawyers relate mostly to international treaties and other binding legal instruments. Since governance by information imposes no legal obligations upon states or individuals, it escapes the established perspective of international lawyers, just as many other instruments, actors, and processes of global governance do. In this part, we propose an instrumental approach for the legal conceptualization of new forms of governance.

Standard instruments as the key category

Instruments as a doctrinal category have a great tradition in continental administrative law for the legal framing of public authority.¹¹⁴ They serve two purposes: on the domestic level, they identify and classify acts of public authority out of the mass of communications emanating from governments and administrations every day; on the international level, there is also a need for such a classificatory function. While the triad of the sources of international law established by Article 38 of the Statute of the International Court of Justice enables a basic classification of the most important binding legal acts, a similar tool is lacking for soft law, as well as for all sorts of non-deontic information, such as governance by information. If one accepts that the concept of international public authority is more comprehensive than the concept of international law as traditionally understood (and as we said, we believe that there are good reasons for doing so), the need arises to classify new forms of public authority such as governance by information.¹¹⁵

Classification is, however, not an end in itself. Rather, it is a precondition for the second function of an instrumental approach: each category of instruments is subject to one identical set of legal requirements. These legal requirements set standards for issues such as the mandate, the procedure, or the formalities necessary for the adoption of such an instrument, and define the outer limits of the content or legal effects such an instrument may have. These legal requirements need to be designed in a way that ensures the legitimacy of the actual instruments.¹¹⁶ As doctrinal categories, standard instruments will eventually become part of the

¹¹⁴ See von Bogdandy and Goldmann, "The Exercise of Public Authority through National Policy Assessment" (n. 97), 269–78.

¹¹⁵ José E. Alvarez, *International Organizations as Law-Makers* (Oxford: Oxford University Press, 2005), 258–9; Matthias Goldmann, "Inside Relative Normativity: From Sources to Standard Instruments for the Exercise of International Public Authority," *German Law Journal* 9 (2008): 1865, 1868–9.

¹¹⁶ On these two functions see, in more detail, von Bogdandy and Goldmann, "The Exercise of Public Authority through National Policy Assessment" (n. 97), 281ff.

existing law.¹¹⁷ Their legal basis resides in the institutional law of a particular international institution, or in general international law.¹¹⁸

What is the advantage of this focus on authoritative instruments? First, it places the question of public authority at the center of public law analysis. Since public law is all about constituting and justifying public authority, we believe that this move is meaningful. Second, an instrumental approach is somewhere in the middle of the road between an approach that focuses on rather abstract (constitutional) principles for the justification of authority, and a case-by-case approach. Although principles provide necessary guidance for the application and further development of the law, they need to be refined by an instrumental optic. Otherwise, their scope of application remains fuzzy, or they might turn out to be over-inclusive with respect to less severe types of public authority, or under-inclusive with respect to more severe types of authority. A case-by-case analysis and critique of new forms of public authority, on the other hand, would hardly be able to provide legitimacy to international public authority. Under any understanding of the rule of law and democratic governance, like cases need to be treated alike, and decision-makers need to find orientation in preceding legal practice. Both require a certain level of abstraction in legal analysis.

Defining a standard instrument: National policy assessment

The definition of standard instruments rests on three points. First, a standard instrument needs to be defined so as to allow for the simple classification of a specific instrument. Instrument-appliers need to be able to classify the instrument they want to apply. Similarly, those who adopt the instrument need to know *ex ante* how it will be classified in order to be able to meet all material and procedural requirements of the specific type of instrument they are going to adopt. Such an *ex ante* perspective is only possible if the parameters used for the definition are formal ones.¹¹⁹ By formal parameters, we understand all those parameters which do not relate to the actual effects of an instrument, the social or normative “ripples”¹²⁰ which it actually generates in the world.

Secondly, even though the definition of a standard instrument is restricted to formal parameters, these formal parameters need to single out those factors which make the instrument authoritative—i.e., those factors that, in an abstract, context-insensitive perspective, can be deemed responsible for the specific impact of the instrument on the social world. Linking legal form and the social world in this way avoids the perils of a power-insensitive positivism and ensures that the standard

¹¹⁷ See above.

¹¹⁸ See in more detail von Bogdandy and Goldmann “The Exercise of Public Authority through National Policy Assessment” (n 97), 292–4.

¹¹⁹ For a more detailed argument in favour of formal criteria see Jan Klabbbers, “Reflections on Soft International Law in a Privatized World,” *Finnish Yearbook of International Law* XVI (2005 (2008)): 311, 322; Jean d’Aspremont, *Formalism and the Sources of International Law* (Oxford: Oxford University Press, 2011), 29ff.; Goldmann, “Inside Relative Normativity” (n.115), 1879.

¹²⁰ Alvarez, *International Organizations* (n. 115), 122, 192.

instrument comprises instruments that, in all probability, constitute an exercise of public authority.

Thirdly, the definition of a standard instrument should be broad enough to allow for meaningful abstraction, and narrow enough to apply the same legal requirements to all instruments falling under the definition. Legal doctrine needs to adopt a normative perspective on the social world and make abstractions rather than looking only at the actual context of a specific act of public authority. However, overly general abstractions are not of much help for solving individual cases. Taking all these criteria together, standard instruments should be defined by means of formal criteria, which suggest that instruments satisfying them typically have a considerable and comparable impact on society and should therefore be subject to the same legal regime.

We now try to put this theory into practice. Using the example of PISA, we propose a definition of a standard instrument called "National Policy Assessment" (NPA). The definition of this standard instrument should be chosen in order for it to cover a wide spectrum of authoritative indicators and instruments aimed at governance by information. Given the very different functions and forms of information promulgated by international institutions,¹²¹ we do not think that one standard instrument necessarily needs to be suitable for all instruments containing information. On the other hand, NPA does not need to be limited to empirical data surveys or (numerical) indicators only.

Which formal characteristics of PISA can be deemed relevant for its actual impact and should be taken as the defining elements of NPA? Of course, a lot of the factors which account for the impact of PISA are context-sensitive and cannot be generalized. Timing and coincidence definitely play a role. PISA has been influential not least because it has addressed a sensitive area of national policy at the right time and in an unprecedented manner.¹²² And of course, the "PISA shock" took place in Germany not least because PISA disappointed previous expectations about the quality of German school education. But we think that some factors are less context-sensitive. We take them as the defining elements of NPA.¹²³

The first element is the fact that PISA produces *empirical information, which claims objectivity*. Of course it is well known that statistical information always involves a degree of contingency, which might stem from the concepts used or from the way in which a survey is carried out. Nevertheless, the results of empirical

¹²¹ See above.

¹²² Markku Lehtonen, "OECD Benchmarking in Enhancing Policy Convergence: Harmonisation, Imposition and Diffusion through the Environmental Performance Reviews?," Conference on International Organizations and Global Environmental Governance, Freie Universität Berlin (2005), 15: Lehtonen argues that PISA is more successful than the OECD Environmental Performance Review, because the OECD has been an "early mover" in educational policy promoting policy innovation, while environmental standards had been agreed in different fora before the OECD started its activities.

¹²³ Although the defining elements of National Policy Assessment are the same as those comprised in our initial definition (cf. von Bogdandy and Goldmann, "The Exercise of Public Authority through National Policy Assessment" (n. 97)) the following refines or modifies the considerations underlying these elements in some respects.

surveys usually claim to come as close to the truth as possible under the current state of the art. This is what gives the information its specific communicative power. It can only be defeated by claims that it is incorrect or inaccurate, not by normative claims. This element distinguishes PISA from compliance reviews that are more normative since they measure the legality (or conformity) of state behavior with a pre-existing treaty or standard.

A second, but related element that makes PISA so specific and influential is the fact that the empirical, apparently objective information that the assessments reveal refers to *policy outcomes*. PISA establishes a relationship between a highly complex social reality and survey data. Almost by necessity, the data only reflects social reality in a simplified, aggregate manner.¹²⁴ This reductionism is characteristic not only of (numerical) indicators, but also of policy assessments that seek to explain social reality in a more qualitative manner. This element distinguishes instruments like PISA from mere fact-finding, which refrains from establishing a nexus with policy.

Besides the claim to objectivity, another factor reinforcing the credibility of PISA is the fact that the entity carrying out the survey and producing the information is different from the entity under scrutiny. Such *external evaluations* of the performance of *domestic policy by international institutions*¹²⁵ might be more immune to manipulation or political abuse, even if produced *for the purposes of the governments under scrutiny*. This element sets National Policy Assessments apart in two further respects:¹²⁶ first, from indicators that serve as internal management tools for international institutions,¹²⁷ and second, from indicators that are primarily designed to suit the purposes of third parties, not those of the evaluated entities.¹²⁸

Two additional features of PISA can be singled out, both of which provide strong incentives for policymakers to take the PISA results seriously. These two features can be summarized in one further defining element, since they jointly contribute to what could be called a *light enforcement mechanism*. The first of these features is the repetitive nature of the PISA tests. From the beginning, PISA was planned to comprise three cycles. Since then, it has been prolonged.¹²⁹ Each subsequent round of PISA holds policymakers accountable for their reactions (or failure to react) to the results of the previous round. This makes it difficult for national policymakers to ignore the PISA tests. Therefore, assessments can be expected to be particularly

¹²⁴ Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 3).

¹²⁵ On our concept of international institution see von Bogdandy, Dann, and Goldmann, "Developing the Publicness of International Public Law" (n. 68).

¹²⁶ These differences are emphasized by Sabino Cassese and Lorenzo Casini, "Public Regulation of Global Indicators," in this volume, pp. 465–74. Davis, Kingsbury, and Merry, "Global Governance by Indicators," this volume, pp. 3–28, also stress the evaluative uses of indicators, but prefer a broader approach.

¹²⁷ Sabino Cassese and Lorenzo Casini, "Public Regulation of Global Indicators," this volume, pp. 465–474 (mentioning the key performance indicators of the Global Fund to Fight AIDS).

¹²⁸ E.g., the country risk indicators produced within the framework of the OECD Export Credits Arrangement, July 21, 2009, TAD/PG(2009).

¹²⁹ Longer Term Strategy of the Development of PISA, 20th meeting of the PISA Governing Board, September 27, 2005, EDU/PISA/GB(2005)21.

relevant if they are repeated, and in particular if carried out by an institution with the reputation of the OECD.¹³⁰

Secondly, the comparative nature of the assessment gives PISA particular bite.¹³¹ Although it might not have been the intention of the initiators of PISA, country rankings appeal to the patriotism of the population and are a strong selling point for the media. Global competition increases the interest of national stakeholders in preventing their country from lagging behind in education. Thus, governments will seek to avoid being named and shamed for less competitive policies, or will react to unfavourable results. Other means, such as specific policy recommendations, might have comparable effects and should be put on par with country rankings in order not to define National Policy Assessments too narrowly.

Thus, National Policy Assessments can be defined as the *revelation of empirical information with a claim to objectivity by international institutions that evaluate the outcomes of domestic policy*, produced *for the purposes of the latter* and coupled with a *light enforcement mechanism* for future domestic policy that relies on the incentives created by iterative evaluations, public disclosure, country rankings, and/or specific policy recommendations.

The legal regime of national policy assessment: Addressing legitimacy concerns

The legal regime applicable to National Policy Assessments needs to be conceptualized so as to ensure the legitimacy of these acts of public authority. This is a task for a dialectical process that includes an analysis of the legitimacy of the instruments falling under the definition of NPA. We will carry this out in the following using the example of PISA. To the extent that these issues reflect structural legitimacy problems of National Policy Assessments, they need to be addressed in the legal regime that we propose.

Before looking into possible legitimacy problems of PISA, we first address why legitimacy is indeed an issue, even if the authors of PISA can truly claim that they carried out a survey of exceptionally high scientific quality.

Is legitimacy an issue? On the contingency of indicators

One might be tempted to consider legitimacy as a non-issue, since PISA produces scientific knowledge, and since knowledge does not need to be justified other than by standards of truth. However, such an argument would ignore the fact that, first, PISA is carried out in an official framework which lends it authority, and second, that it involves important policy choices.

Regarding the first point, we see a fundamental difference between NPAs like PISA, which are adopted and carried out by governments and international

¹³⁰ On the importance of reputation, see Jason C. Sharman, "Rationalist and Constructivist Perspectives on Reputation," *Political Studies* 55 (2007): 20.

¹³¹ Lehtonen, "OECD Benchmarking in Enhancing Policy Convergence" (n. 122), 15; Martens, "Comparing Governance of International Organisations" (n. 92).

institutions in the frame of their competence, and private surveys. Scholars or journalists acting in their private capacity may of course invoke fundamental rights, notably the right to free speech, no matter whether their research is funded by public or private sources. Some constitutions even recognize academic freedom as a fundamental right of its own (e.g., Article 5 para. 3 of the German Basic Law, or Article 14 s. 5 para. 2 of the 1987 Philippine Constitution). By contrast, the PISA reports are official OECD documents, endorsed by the PGB and therefore endowed with the authority of the institution. Those who yield public authority may not invoke liberty or their fundamental rights without restrictions. Rather, fundamental rights safeguard private rights and institutions *against* public authority.¹³² The exercise of public authority and the exercise of an individual right therefore follow opposite logics.

Regarding the second point, the PISA reports are based on a state-of-the-art survey. Designers of the survey invested significant energy in producing a culturally neutral survey. It favors measurement of general competencies instead of specific knowledge, which hardly would have done justice to the different curricula used in the participating states. Nevertheless, the design and use of an indicator involves contingencies that necessitate a legitimating legal framework. An indicator is based on the hypothesis that there is a causal relationship between some statistical value (e.g., student–teacher ratio) and a social phenomenon (e.g., the quality of school education). Both the hypothesis and the concepts to which it refers can be highly contingent. The hypothesis might be corroborated by statistical correlation, but this is not the same as causality, since other relevant factors may play a role.¹³³ Concepts only have a mental existence. They are created by an act of definition, not one of recognition. This conceptual contingency is best illustrated by the concept of education used for PISA. As has been shown above, it is a highly contingent question whether output-oriented indicators can be or should be used in order to assess the quality of school education.¹³⁴ Such contingencies create the need for a legitimating legal framework, even if they are implied in any indicator.

Input legitimacy: PISA as an example of bureaucracy-driven governance

Input legitimacy refers to ascending modes of legitimizing acts of public authority, such as authorization, procedures, and review. In order to assess the input legitimacy of PISA, it is important not only to look at formal criteria, but also to analyze how relevant decisions actually have been taken.

At the outset, the activities of the OECD in the field of school education were rooted in the desire of the US and French governments to play the puck against the board and create pressure on the international level in order to advance domestic policies. However, in the end, the activities of the OECD developed an entirely

¹³² E.g., Article 1 of the European Convention for the Protection of Human Rights and Fundamental Freedoms: “The High Contracting Parties shall secure to everyone within their jurisdiction the rights and freedoms defined in Section I of this Convention.”

¹³³ Instructive: Desrosières, *The Politics of Large Numbers* (n. 1), 105ff.

¹³⁴ See above.

different dynamic. Instead of national executives, expert-driven transnational bureaucracies took the lead, achieved the adoption of PISA, and played a decisive role in its implementation.¹³⁵

Regarding the adoption of PISA, it was indeed the influence of a transnational bureaucracy which changed the opinion of OECD member states in favor of PISA. While member states were mostly skeptical towards PISA in 1995, they adopted it only two years later.¹³⁶ How did this change of minds among the OECD member states come about within only two years? Of course, different factors contributed to this result. First, empirical student assessments became more common, more policy relevant and of better quality during the 1990s.¹³⁷ In particular the assessments carried out by the IEA paved the way for PISA. Because states had some prior experience with empirical assessments of educational outcomes, subscribing to PISA was no longer as much of an excursion into unknown deep waters.

However, this alone does not explain why PISA came to be adopted. To a considerable extent, this goes back to the influence of interested people in national administrations and, above all, in the OECD Secretariat, as well as of private lobbyists. Unlike national governments, the OECD Secretariat became much more determined to set up PISA, after initial resistance. Two factors contributed to this. First, the OECD Secretariat had moved towards an economic, output-oriented concept of education. This move resulted from a general ideological shift in the OECD Secretariat. In the 1960s and 1970s, the OECD Secretariat had been dominated by Keynesian economics. Since the 1980s, as a consequence of the policies of the Reagan and Thatcher governments, the Secretariat increasingly subscribed to a supply-oriented, neoliberal economic philosophy.¹³⁸ This also had an impact on its educational policy. It began considering education as just another sector of welfare state administration in need of reform.¹³⁹ Thus, while the OECD Secretariat had been all but enthusiastic to work on educational indicators in the late 1980s, by the middle of the 1990s, it considered educational indicators as a valuable policy tool for ensuring a high level of education that could stimulate economic supply. The OECD Secretariat became a promoter of the same intellectual shift that had previously reshaped the Secretariat itself.¹⁴⁰

¹³⁵ Martens and Wolf, "Paradoxien der Neuen Staatsräson" (n. 9).

¹³⁶ See above.

¹³⁷ For an overview see Eugene Owen, Maria Stephens, Jay Moskowitz, and Guillermo Gil, "From 'Horse Race' to Educational Improvement: The Future of International Education Assessments," in INES Compendium (n. 25), 7–18, paras 8–13.

¹³⁸ Martin Marcussen, "The OECD in Search of a Role: Playing the Idea Game," European Consortium for Political Research (2001), available at <<http://www.essex.ac.uk/ecpr/events/jointsessions/paperarchive/grenoble/ws12/marcussen.pdf>>, 8; Henry et al., *The OECD, Globalization and Education Policy* (n. 111), 33–7; Noaksson and Jacobsson, "The Production of Ideas and Expert Knowledge in OECD" (n. 92), 48.

¹³⁹ Henry et al., *The OECD, Globalisation and Education Policy* (n. 111), 33.

¹⁴⁰ Ibid., 37.

Secondly, apart from its ideological predisposition, the OECD also had a very concrete strategic interest in carrying out projects such as PISA. The end of the Cold War had brought the OECD into crisis. Governments did not see a need for an expensive “think-tank of the West,” as the concept of the West had lost its previous significance. The OECD therefore had a strong interest in carrying out innovative projects that governments would consider useful.¹⁴¹

Motivated by these considerations, the Secretariat and like-minded government representatives set up a process that one might call the “bureaucratic entanglement” of national governments. Despite the mixed reactions of member states in 1995, in the INES Networks, the Secretariat as well as like-minded states representatives simply continued their work on the data strategy on an expert level. Representatives of other states did not object, but took a more or less active part in the development of the strategy. Overall political concerns, such as the significance of such a project for the approach to education, the ratio between the costs and the political benefit of the survey data, and its impact on state sovereignty, did not seem as salient at the expert level.

Once the expert level had developed and agreed on the outlines of the data strategy, however, it became difficult for individual states to stop the entire process on the supervisory, politically responsible level such as the Education Committee. This has to do with the particular relationships between supervisory and subsidiary bodies in the OECD, like the Education Committee and INES. It would be inaccurate to imagine this relationship as a one-way street between a principal and an agent. Rather, as described by one interview partner, it is an interactive relationship oscillating between top-down and bottom-up processes, in which the different levels mutually influence each other. Actors on the upper level depend on the lower levels elaborating proposals according to their directives, and lower levels depend on the periodic approval of their work by the upper levels. If the lower level develops a concise proposal with the support, or at least the tacit consent, of the upper level, when subsequently taking a decision on the proposal, the upper level can hardly ignore its previous action or tacit consent. Conversely, if a state has fundamental concerns with a project, it is expected to articulate these at any level before the proposal is spelled out in all details. Otherwise, a complex organization of formally equal members with a plethora of specialized committees would be inoperable. As a result, those members of the OECD who had not been enthusiastic for the project in 1995 found themselves increasingly entrapped by a process that had developed a dynamic of its own.¹⁴²

The lobbying efforts of companies with a commercial interest in PISA facilitated this strategy.¹⁴³ Recent decades have seen the emergence of a market for educational assessments in which commercial institutions such as the Australian Council

¹⁴¹ Cf. OECD, *The OECD—Challenges and Strategic Objectives: 1997*. Note by the Secretary General, C(97)180, September 10, 1997.

¹⁴² Martens and Wolf, “Paradoxien der Neuen Staatsräson” (n. 9), 167–9.

¹⁴³ Elisabeth Flitner, “Pädagogische Wertschöpfung. Zur Rationalisierung von Schulsystemen durch public-private-partnerships am Beispiel von PISA,” in Jürgen Oelkers et al. (eds), *Rationalität und Bildung. Studien im Umkreis Max Webers* (2006), 245, 246.

on Educational Research (which became the main international contractor for PISA) and the US-based Educational Testing Service and Westat companies offer their services to governments.

After adoption, bureaucratic entrapment has also played a role in the implementation of PISA. Even though the PGB has a very strong position within the PISA framework, it would be a deception to see its decisions as an ideal type of state consent. Rather, the control of the participating states over the design of the PISA and dissemination of its results through the PGB has some factual limits. The implementation of PISA has been an unprecedented effort requiring a very high level of expertise. It has been steered by a closely interlinked epistemic community consisting of government experts, independent experts, Secretariat staff, staff of the PISA Consortium, and national contractors, some of whom fill several functions at once. Many of them already worked together in INES or other venues before the birth of PISA. Their expertise gives them considerable leeway to act in accordance with their own convictions and provides some shelter against high-level government instructions. Even the role of the PGB was sometimes more reactive than proactive. The busy schedules of its sessions left no space for the development of substantive proposals, which had to be entrusted to expert groups, the Consortium, or the Secretariat. In addition, the entire process became so complex that, per our interviewees, few people outside the OECD Secretariat fully understand its procedural aspects. “Ask Andreas” was a frequent answer to our questions.¹⁴⁴ The strong position of the Secretariat might also be due to the continuity of its key personnel.

In conclusion, bureaucracies and their dynamics have heavily influenced both the adoption and the implementation of PISA. Even though most of the participants in this process enjoy some personal legitimacy, having been appointed by their governments or by the OECD Secretary General, this personal legitimacy alone hardly suffices to justify the dramatic policy changes triggered by PISA.

Consequences for the legal regime of NPAs

The problematic input legitimacy of PISA could indeed be a structural problem of NPAs. Since empirical assessments are a very difficult matter, requiring a high level of expertise and very often involving commercial interests, there is reason to believe that similar difficulties might arise (or have arisen) with other National Policy Assessments. Therefore, the legal regime applicable to NPAs should address these questions. In this respect, we propose three principles for the legal regime of PISA. They relate to the mandate, the adoption procedure, and the implementation of NPAs, respectively. The fact that these candidates have to varying extents already been included in the present legal framework of PISA corroborates our view that these issues are indeed crucial for the legitimacy of PISA.

¹⁴⁴ Referring to Andreas Schleicher, Head of the Indicators and Analysis Division, OECD Directorate for Education.

First, regarding the mandate, we believe that a precise mandate outlining not only the costs, but also the content and political direction of the envisaged project is a necessary element of NPAs. Such a mandate is not required under the law of international organizations in its present state, especially since PISA does not amount to the enactment of a binding legal instrument.¹⁴⁵ This is a deficit that should be corrected. The need for a legal basis is one of the most fundamental means by which public law provides legitimacy.¹⁴⁶ The legal basis should be precise enough to ensure that the main political decisions remain in the hands of politically visible and accountable bodies such as the OECD Council. Basing an NPA on the general competence of the Secretariat under the founding treaty of an international institution would not be commensurate with its political potential. For example, Article 1 of the OECD Convention, which outlines the aims of the organization, is overly broad. Almost any possible NPA would fall under it. PISA is already based on a clear and exhaustive legal mandate. The Strategy document endorsed by the OECD Council lays down the essential features of the policy. Even though the indicators were not defined in this document but only later during the implementation stage, the Strategy describes in abstract ways how the indicators should be designed and therefore determines the overall policy thrust of PISA.

Second, the adoption process of NPAs should be designed in a way that enables a political debate involving many stakeholders on the national and international levels prior to adoption. Such a requirement would reduce the risk of back-room expert discussions leading to bureaucratic entanglement. This requirement could be fulfilled in different ways. For example, the two-staged adoption process including an exchange of notes with governments as well as the adoption of the project by the OECD Council could be used in order to trigger a political debate on the national level. This process should include consultations with domestic stakeholders. Teachers, researchers, parent representatives, and politicians should debate the design of tests and reflect on the concept of education proposed. Such efforts might increase the legitimacy of an NPA and even be considered an adequate remedy for the lack of parliamentary ratification of the decision establishing PISA. It should therefore be taken as an essential element of the legal framework of NPAs.

Third, the political discretion exercised in the implementation of NPAs on the international level requires additional safeguards to ensure the legitimacy of the process. In particular, it requires that all decisions regarding the design of indicators should be taken by a body composed of state representatives, like the PGB in case of PISA.

Nevertheless, in many instances, the bureaucratic apparatus will have to rely on external expertise in order to implement an assessment. The use of external experts might also provide a safeguard against structural biases in the Secretariat of the

¹⁴⁵ In detail Jan Klabbers, *An Introduction to International Institutional Law* (Cambridge: Cambridge University Press, 2002), 60ff.

¹⁴⁶ For a comparative study, see Armin von Bogdandy, *Gubernative Rechtsetzung* (Tübingen: Mohr Siebeck, 2000), 166ff.

organization carrying out the assessment. Hence it appears to be a necessary element of the legal regime of NPAs. In the case of PISA, this requirement was fulfilled, as shown by the involvement of experts in the field in the functional expert groups of PISA.¹⁴⁷ On the other hand, it appears necessary to ensure a geographically balanced selection of experts in order to avoid national or regional biases. In this respect, it is remarkable that the members of the PISA functional expert groups were selected not only on the basis of their academic qualifications, but also of their national origin. In fact, the Strategy encourages a balanced country representation in the expert groups and envisages that states not represented in a particular expert group charge another expert with their representation.¹⁴⁸ This shows that the drafters of PISA recognized the contingency of expert advice. However, in addition to the attention paid to geographic representation, it might be advisable to issue an open call for applications in order to exclude any bias in the selection of experts.

Output legitimacy: Evidence-based policymaking or policy-based evidence-making?

By output legitimacy, we understand a descending mode of legitimizing public authority: it analyzes whether an act of public authority reaches the goal for which it was created.¹⁴⁹ PISA aims at measuring the outcomes of education and providing policymakers with relevant information on the performance of their schools. It can be understood as a contribution to a deliberative model of democracy that enables policymakers to make well-informed decisions. This requires that the information provided be trustworthy, sufficiently accurate, and not unduly influenced by the interests of those whose policies are being assessed. Otherwise, NPAs would fail to link society with policymaking, which is the very point of a deliberative model of democracy.¹⁵⁰ The question therefore is whether PISA meets this standard. Two different kinds of doubts can be raised, relating to the scientific quality and the political neutrality of the study, respectively.

First, in spite of the efforts of those charged with the implementation of PISA to make the study as accurate as possible, numerous doubts have been raised about the scientific quality of PISA, in particular the methods used and the interpretation of the results. While it is readily apparent that much of the public critique consists of populist statements by interested persons, the scholarly critique cannot be discarded that easily. Some analysts contend that PISA does not measure educational achievements, but resembles an intelligence test.¹⁵¹ Others question the composi-

¹⁴⁷ See above.

¹⁴⁸ DEELSA/ED/CERI/CD(97)7, August 19, 1997, para. 80.

¹⁴⁹ Fritz W. Scharpf, *Regieren in Europa* (Frankfurt a. M.: Campus, 1999), 20ff.

¹⁵⁰ Cf. Habermas, *Faktizität und Geltung* (n. 89), 361–6.

¹⁵¹ Heiner Rindermann, "Was messen internationale Schulleistungstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz?" *Psychologische Rundschau* 57 (2006): 69; see also the replies by Manfred Prenzel, Oliver Walter, and Andreas Frey, "PISA misst Kompetenzen," *Psychologische Rundschau* 57 (2006): 128; Jürgen Baumert et al., "Was messen internationale Schulleistungstudien?—Resultate kumulativer Wissenserwerbsprozesse," *Psychologische Rundschau* 58 (2007): 118.

tion of the so-called risk group, i.e., of students who do not meet certain minimum requirements.¹⁵² Both critiques have major consequences for the interpretation of the test results.

Other scholars criticize the test design and implementation. For example, it has been argued that students used to multiple-choice tests might attain better results than students from countries where multiple-choice tests are not common. Also, the examples of the test questions that were published evoked much critique. Some find the answers and their evaluation questionable. In particular, the use of translated texts might compromise the comparability of the test, as translations often tend to be interpretative, longer than the original, and prone to mistakes.¹⁵³ Prais looks at discrepancies between the good British results in the math section of PISA and the lower results in TIMSS and explains these in particular by the low British response rate to PISA.¹⁵⁴

This is not the place for an in-depth engagement with this criticism. Rather, at this stage we ask how this affects the output legitimacy of PISA. From a deliberative perspective, one could argue that PISA is just one contribution in a communicative process, and that criticism, even fundamental criticism, is just a normal consequence of any scientific publication. However, PISA is not only science, it is also politics, and it is authoritative. Therefore we think that some mechanisms need to be put in place in order to take these criticisms into account in the making of PISA.

A second kind of criticism relates to the political influence exercised on what purports to be an empirical survey. Of course, PISA needed some political guidance during the implementation stage, as the design of the indicators is politically highly sensitive. However, it is questionable how much influence should be afforded to national governments during the drafting stage of the PISA reports and during release. For example, the PGB pressed several times for policy relevance in the reports.¹⁵⁵ In at least one case, this led to a last-minute change to the PISA reports by the Secretariat. This was even criticized by scholars involved in PISA, who held that the released data could not be verified.¹⁵⁶

The risk of political capture became evident when some German education ministers disagreed with the OECD Secretariat about the evaluation of the PISA results. After the 2006 cycle, German education ministers wanted to derive a positive result from the trend indicators, which in their opinion showed that the performance of German students had improved. The OECD Secretariat expressed doubts about this interpretation, pointing to differences in the assessment designs

¹⁵² E.g., Klaus Klemm, "Schulforscher, aufgepasst!" *Die Zeit*, May 15, 2008, available at <<http://www.zeit.de/2008/21/C-Bildungsforschung>>.

¹⁵³ For these and many more critiques see Thomas Jahnke and Wolfram Meyerhöfer (eds), *PISA & Co.—Kritik eines Programms* (Hildesheim: Verlag Franzbecker, 2007).

¹⁵⁴ S. J. Prais, "Cautions on OECD'S Recent Educational Survey (PISA)," *Oxford Review of Education* 29 (2003): 139.

¹⁵⁵ E.g., Seventh Meeting of the BPC, October 4–5, 1999, DEELSA/PISA/BPC/M(99)1, 3–4; Tenth Meeting of the BPC, April 18–20, 2001, DEELSA/PISA/BPC/M(2001)1, 7.

¹⁵⁶ Manfred Prenzel, Jürgen Baumert, and Ekkehard Klieme, "Falscher Verdacht," *Die Zeit*, May 29, 2008, available at <http://blog.bildungsserver.de/wp-content/uploads/2008/05/replikklemm_vollversion.pdf>, 73–4.

between the previous assessment and the 2006 data. German education ministers, as well as the German National PISA Coordinator, strongly disagreed. Some education ministers even threatened to discontinue their participation in PISA if the OECD Secretariat did not adopt their interpretation of the results.¹⁵⁷ The damage of such an argument to the deliberative value of PISA is evident.

Consequences for the legal regime of NPAs

Both the treatment of criticism and the risk of capture are structural problems that may arise with any NPA and need to be addressed by the legal regime of NPAs. Of course, the legal regime of NPAs cannot ensure that the results of empirical surveys reflect “truth” in the true sense of the word. If truth is defined as the latest state of error, such an attempt would be futile. However, it should define standards which are commensurate with the fact that NPAs are not just some kind of communication contributing to a public debate, but that they are an authoritative, politically produced statement. In other words, the legal regime needs to ensure that NPAs do provide a link between society and policymaking (which is the point of a deliberative democracy) and do not amount to an uncritical self-affirmation of policymaking by policymakers. In this respect, it goes almost without saying that NPAs should be required to *respect pertinent scientific standards*. Two further aspects need to be taken into account, one of them addressing scholarly criticism, the other, the risk of capture.

Regarding the first point, just as judges are required to give reasons for their decisions, the producers of NPAs also need to give reasons and be reasonably open to criticism. PISA does indeed show a high standard here in some respects. For example, scholars are granted access to the data basis of the assessment, and a detailed technical report is addressed to a professional audience. This enables profound, critical scrutiny of the survey and should be considered a part of the legal regime of NPAs. That most test questions are kept confidential is deplorable, but necessary in order to use them again in subsequent test cycles so that the results from different cycles remain comparable.

However, as has been said, NPAs should take critical voices seriously and provide some mechanism for incorporating feedback. In this respect, PISA could be improved.¹⁵⁸ While some of the scholars involved in PISA did their best to respond to the criticism raised against the test results on a private basis,¹⁵⁹ the official response to such criticism has been assailed as being insufficiently receptive. Another problem is that the critiques, as well as the respective replies, are spread over a myriad of isolated publications, including scientific and not-so-scientific

¹⁵⁷ Cf. Landtag von Baden-Württemberg, Antrag der Fraktion der SPD und Stellungnahme des Ministeriums für Kultus, Jugend und Sport, “Kein Ausstieg Baden-Württembergs aus PISA 2012,” November 3, 2009, Drs. 14/5375, available at <http://www.landtag-bw.de/WP14/Drucksachen/5000/14_5375_d.pdf>, 2–3.

¹⁵⁸ This further specifies the legal regime proposed in von Bogdandy and Goldmann, “The Exercise of Public Authority through National Policy Assessment” (n. 97).

¹⁵⁹ See in particular Prenzel, Baumert and others, “Falscher Verdacht” (n. 156).

ones. The solution might lie in the establishment of an institutionalized forum for debate, for example a roundtable discussion which provides a platform for exchange—something that could be reproduced in the legal regime of other NPAs. The OECD has an excellent dissemination strategy for the media, which includes seminars at which journalists are instructed how to understand the survey. These are laudable efforts, which foster the deliberative goals of PISA. Such efforts could be extended to encompass the official reaction to criticism. They could even take place before the PISA results are published, which would give them the character of a peer-review—a recognized method for ascertaining scientific quality. Confidentiality agreements could ensure that this does not compromise the dissemination strategy.

Regarding the risk of capture, the legal regime of NPAs should also address the roles of governments and the OECD when drafting the reports. A careful analysis of the PISA framework and of the dissemination strategy shows that PISA is thus far dominated by a principle of the national ownership of the test results. For example, the international reports refrain from drawing recommendations from the empirical results; every word needs to be approved by the PGB and each country has the power to stop the release of its data. The participating states thus have the results of the study in their hands to the greatest extent possible.

Of course, governments have every reason to insist on this kind of national ownership given the sensitivity of the issue of school education and the fact that the reports assess a system of education for which they bear responsibility. Nevertheless, national ownership as it stands is not mindful enough of the risk of capture. This might compromise the deliberative character of NPAs. Governments need to accept that they might receive bad results. Their interests are safeguarded by the impact that they have on the test design. Therefore, it would be advisable to give the Secretariat more autonomy in the drafting of the reports.¹⁶⁰ For example, it might be advisable to give governments only a consultative role, and to have the draft reports prepared by the Secretariat and the Consortium, and approved by independent experts in order to verify their accuracy. Governments may still publish their dissenting opinions. This kind of independent review is not so uncommon. For example, the IMF engages in such a practice in its Staff Reports and World Economic Outlooks.¹⁶¹ Conversely, any further autonomy on the part of international bureaucracy beyond that which is required to safeguard against political capture merits criticism.

¹⁶⁰ This repeats the opinion expressed in von Bogdandy and Goldmann, “The Exercise of Public Authority through National Policy Assessment” (n. 97).

¹⁶¹ Schäfer, *Die neue Unverbindlichkeit* (n. 2), 118.

NPAs as the avant-garde of an information-based multilevel democracy?

Our analysis has provided a theoretical framework for the legal analysis of an instrument which bases its authority on governance by information. The example of PISA demonstrates that such instruments need an elaborate legal framework ensuring both the legitimacy and the effectiveness of the public authority exercised in this way. Standard instruments are a useful doctrinal category for establishing a legal regime for indicators of this kind. They provide policymakers in international institutions with the knowledge they need in order to set out such policies in other issue areas. Once a standard instrument has been elaborated, it can be applied to multiple issue areas. This also has a rationalizing effect. Given the plethora of governance fora and instruments that emerged since about the end of the Cold War, something is necessary in order to rationalize and structure the ways in which we are governed today.

Indeed, NPAs seem to be spreading more and more. After the success of PISA, the OECD plans to expand these activities in the educational sector.¹⁶² And the recent global financial crisis has shown that there is an urgent need for external evaluations of government policies, which might address issues like budget deficits or financial market regulation. The IMF's Global Financial Stability Reports are a first step in this direction.

These examples and proposals indicate that NPAs might develop into an integral part of a multilevel system of governance. As performance has an impact on the legitimacy of public authority, it may become a duty for national governments to expose themselves to NPAs. As states are less and less able to meet the needs of a globalized world, and as their citizens and economies must compete in worldwide markets, the legitimacy of the state-based structure of the international legal order might fade if states do not continue to perform at a high level. Experience to date suggests that the German public would take it very badly if their government decided to discontinue its participation in PISA. As globalization proceeds, such expectations may consolidate into legal obligations, just as there is a duty today to enable the participation of each citizen in the conduct of public affairs.¹⁶³

¹⁶² Similar large-scale assessments of adult competencies and the outcomes of higher education are currently being explored by the OECD, for example the envisaged OECD Programme for the International Assessment of Adult Competencies.

¹⁶³ Cf. Art. 25(a) of the International Covenant on Civil and Political Rights.

The Dynamism of Indicators*

Wendy Nelson Espeland and Michael Sauder

Introduction

A common refrain of analysts of social indicators is that indicators are often dynamic, seeming to possess “a life of their own.” That is, even when there are clear ideas about what a particular indicator is intended to do, what it means, or who will use it, it can function in ways and places that seem independent of its origins. This “dynamism” will depend on which groups notice or ignore an indicator, how they use it, whether they become invested in it, and how these groups are connected to one another. In this way, indicators are like other forms of knowledge that become authoritative as they accumulate networks of constituents, technologies, and things.¹

“Successful” indicators—those that become influential and institutionalized—often produce powerful and unanticipated effects: they can change how people think about what they do, what is comparable, how excellence or mediocrity is defined, or even who they are. As the investigations of indicators in this volume demonstrate, indicators are restless signifiers whose consequences and significance can be hard to predict or even contain.² They reconstruct the places and people that they purport to measure,³ change how power and expertise are

* Thanks to Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and especially Sally Engle Merry for encouragement and editorial discernment. Thanks also to those at the NYU conference and the Law and Society Annual Meeting Panel on Indicators and Global Governance organized by Sally Engle Merry, and to Bruce Carruthers and Lauren Buxbaum for helpful comments. This research was supported by the Robert Wood Johnson Foundation Scholars in Health Policy Research Program.

¹ See Bruno Latour, *Science in Action: How to Follow Scientists and Engineers Through Society* (Cambridge: Harvard University Press, 1988); Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford: Oxford University Press, 2007).

² On the unintended consequences of indicators, see Katarina Pistor, “Re-Construction of Private Indicators for Public Purposes,” this volume, pp. 165–79; Galit Safarty, “Perils and Promise of Indicators in Global Governance: A Case Study of Corporate Sustainability Reporting,” paper presented at Law and Society Association Annual Meeting, San Francisco 2011.

³ Terence C. Halliday, “Legal Yardsticks: International Financial Institutions as Diagnosticians and Designers of the Laws of Nations,” this volume, pp. 180–216; Marina Zaloznaya and John Hagen, “Fighting Human Trafficking or Instituting Totalitarian Control? The Political Co-optation of Human Rights Protection in Belarus,” this volume, pp. 344–64.

mobilized,⁴ and often reproduce themselves as they prompt the creation of new layers or tiers of quantified evaluation.⁵

Indicators produce these dynamic effects for many reasons. Indicators simplify complex and evolving institutions and processes, which is one reason why they are so useful. This simplification, however, is often forgotten or ignored when they are put into practice. Collecting the data that comprises indicators is often more difficult and messy than appears from the outside; the phenomena indicators are designed to measure change over time, and indicators may not be revised in ways that reflect these changes. Finally, and most importantly, the people being measured, assessed, and compared often respond to the act of being measured in ways that subvert the intentions that first motivated the assessment process.

This article considers some factors that affect the movements and authority of indicators. In order to understand the dynamic nature of indicators—how they take on new meanings and new uses and how they seem to become autonomous and distant from their origins—it is crucial to learn how such measures are used rather than simply assume that their effects are known. This, in turn, requires research that looks *inside* organizations to learn how indicators are made, interpreted, and circulate among groups of users.

Toward that end, we draw on an intensive investigation of *US News & World Report* (USN) educational rankings, emphasizing law schools, to identify patterns that facilitate the movement and reinterpretation of indicators and to describe some of the conditions that influence or mediate that dynamism. After briefly outlining the development and components of USN rankings and the methods we used to investigate them, we examine three broad processes that shape how people respond to indicators in ways that propel them to new uses and meanings: the commensuration inherent in the creation of rankings; how people react to indicators; and the mediation of the meaning and effectiveness of indicators as they are employed by groups. We next analyze these three characteristics in relation to two prominent global educational rankings: the QS World University Rankings (QSWUR) and Shanghai Jiao Tong Academic Ranking of World Universities (ARWU). We conclude with suggestions about how indicators have shaped the governance of higher education and offer some general lessons for research on indicators more broadly.

⁴ Angelina Fisher, "From Diagnosing Under-immunization to Evaluating Health Care Systems: Immunization Coverage Indicators as Technology of Global Governance," this volume, pp. 217–46; René Urueña, "Internally Displaced Population in Columbia: A Case Study on the Domestic Aspects of Indicators as Technologies of Global Governance," this volume, pp. 249–80; Wendy Espeland and Berit Vannebo, "Accountability, Quantification and Law," *Annual Review of Law and Social Science* 3 (2007): 21–43.

⁵ Ronan Shamir and Dana Weiss, "Semiotics of Indicators: The Case of Corporate Human Rights Responsibility," this volume, pp. 110–31.

The evolution of a “successful” indicator: educational rankings

As the introductory essay to this volume suggests, it is important to conceive of and analyze indicators as particular forms of knowledge produced under circumstances that may shift.⁶ Changing contexts or conditions, the accumulation of new audiences and uses, or even the increasing familiarity or visibility of an indicator may change its meaning. One of the distinctive features of the production of rankings as indicators of educational quality is that they were first produced by a for-profit news magazine with no particular expertise in education. These origins left their mark on rankings as other businesses and organizations appropriated and adapted the magazine’s model.

Contemporary rankings of educational institutions had humble beginnings. Although rankings of universities have existed for about 100 years, early examples were intermittent and prepared mainly for educators. It was not until the 1980s that media began to produce annual rankings of higher educational institutions that were targeted to the general public. The first media ranking of universities was prepared by *US News*, then a weekly news magazine.⁷ In 1983, *USN* conducted a survey of college presidents asking them to rank the best American colleges and universities. The results generated so much interest that the magazine began publishing annual rankings of colleges in 1985. In 1987 it developed similar rankings of graduate and professional schools, including law schools. The rankings issues became best-sellers for *USN*, and this success led to other lucrative endeavors such as the publication of guidebooks for college and graduate applicants and, eventually, the construction of a popular website with detailed ranking information. In 2007, *USN* publisher Kerry Dyer reported that when college rankings were released, within 72 hours its website received 10 million visits and that a typical college ranking issue sold 45,000 copies on a newsstand, 50 percent more than a regular issue.⁸ As one former *USN* editor, Mel Elfin, put it, rankings “saved the franchise.”⁹

When Mort Zuckerman, a wealthy real estate investor and owner of the *New York Daily News*, took over the editorship of *USN* in 1983, he was determined to raise its profile by establishing a distinctive and appealing identity for a periodical that was for many people an afterthought to the more popular *Time* and *Newsweek*

⁶ Kevin Davis, Benedict Kingsbury, and Sally Engle Merry, “Introduction: Global Governance by Indicators,” this volume, pp. 3–28.

⁷ For a more detailed history on the origins of the *USN* rankings, see David S. Webster, “Rankings of Undergraduate Education in *US News & World Report* and Money: Are They Any Good?” *Change* 24 (1992): 18–31; Wendy Espeland and Michael Sauder, “Rankings and Reactivity: How Public Measures Recreate Social Worlds,” *American Journal of Sociology* 113 (2007): 1–30; Dorie Bertram, “Annotated Bibliography: Ranking of Law Schools by *US News & World Report*,” Legal Studies Research Paper Series, 10-08-03, Washington University in St Louis, August 13, 2010.

⁸ Samuel G. Freedman, “Putting a Curious Eye on a High School Ranking System,” *New York Times*, 2007.

⁹ Roger Parloff, “Who’s Number One? And Who’s Number 52, 91, and 137?,” *American Lawyer*, April 1998.

magazines. Zuckerman's emphasis would become "news you can use," with rankings as an exemplar of this new orientation. Much like *Consumer Reports*, a magazine devoted to evaluating products, rankings were intended to help prospective students make decisions about where to apply and attend college and graduate school. Writing about law school rankings, Robert Morse, who oversees the production of all rankings at *USN* says:

The main purpose of the rankings is to provide prospective law school students with much-needed—and clearly desired—comparative information to help them make decisions on where to apply and enroll. In today's legal job market a student's choice of law school plays a considerable role in getting that all-important first legal job. That job is particularly important since some new law school graduates have accumulated over \$150,000 in debt just to get their J.D. degree and many need to start paying off their student loans.¹⁰

In short, *USN* sees its rankings as useful measures that provide clear, objective, comparative information about school quality.

Like many indicators, rankings do much more than reflect existing differences in law schools. They also produce important changes in how schools and applicants behave, effectively transforming the objects they measure. Law school rankings have influenced almost every aspect of legal education, despite the fact that they were initially denounced by most administrators and by virtually all of the professional organizations associated with legal education. They have changed which schools students apply to and attend. They have changed how resources are distributed within the schools; for example, schools have shifted money from need-based to merit-based scholarships in order to raise their median LSAT scores and improve their rank. Rankings have also changed the work routines and career paths of many administrators; career service personnel, for instance, now spend much more time tracking down students to find out if they are employed—a key component of the ranking formula—often at the expense of additional individual career counseling or networking with employers. Rankings encourage conformity among law schools in that law schools devoted to missions not captured by *USN*'s formula feel pressure to change those missions.¹¹

So why have these controversial, catalyzing, and now ubiquitous measures become so consequential for legal education? What can rankings teach us about indicators and regulation more generally? Before describing how *USN* rankings produce these effects, it is useful to know more about how they are made.

The algorithm *USN* uses to create rankings includes four main categories: reputation, selectivity, placement, and faculty resources. Performance is determined by a collection of weighted measures within that category. Selectivity, for example, determines 25 percent of the overall rank and is based on three components:

¹⁰ <<http://www.usnews.com/blogs/college-rankings-blog/2009/02/02/do-the-rankings-punish-law-schools.html>>.

¹¹ These effects of rankings are discussed in greater detail in Michael Sauder and Ryon Lancaster, "Do Rankings Matter? The Effects of *US News & World Report* Rankings on the Admission Process of Law Schools," *Law and Society Review* 40 (2006): 105–34; Espeland and Sauder, "Rankings and Reactivity" (n. 7); Michael Sauder and Wendy Espeland, "The Discipline of Rankings: Tight Coupling and Organizational Change," *American Sociological Review* 74 (2009): 63–82.

student LSAT scores count for 50 percent of the selectivity ranking (12.5 percent of the overall score); GPA represents 40 percent (10 percent overall); and acceptance rate accounts for 10 percent of the selectivity score (2.5 percent overall). The scores from these four main categories of criteria are summed to create a composite score that is scaled to create a school's overall rank.¹²

The success of *USN* educational rankings of educational quality prompted first magazines and newspapers (e.g., *Times Educational Supplement*, *Macleans*, *Der Spiegel*, *Newsweek*, *Asiaweek*), then data collection and consulting firms (Quacquarelli Symonds Limited, Reuters), and then national and international governmental organizations (the Organisation for Economic Cooperation and Development's PISA rankings)¹³ to promote their own versions of educational rankings. Within *USN*, rankings have undergone a gradual elaboration of their scope and presentation, a pattern that also characterizes many global indicators. Most importantly, the rankings have expanded their coverage and have become increasingly precise over time. The first rankings of law schools (1987–1990) included only the top 25 schools; *USN* gradually expanded to include all schools accredited by the American Bar Association (now about 190), dividing schools not ranked in the top 50 into three hierarchical tiers (2nd, 3rd, 4th tier) in which schools were listed alphabetically. *USN* expanded its ordinal list to include, first, the top 100 schools, and in 2005, the top 150 schools, presenting the remaining schools in hierarchical tiers (first 3rd and 4th tier, now just the 4th tier).¹⁴

It is important to note that the effects of indicators are not restricted to their targets. Many national and international magazines took note of *USN*'s success with rankings and devised their own educational rankings, often relying on statistics that are readily available or easy to compile. Like a lot of magazines, *USN* has recently struggled with declining circulation and advertising revenues. In 2008, the magazine scaled back first from a weekly to a biweekly and then to a monthly magazine. *USN* printed its last issue for subscribers in December 2010. Instead of a monthly news publication, *USN* offers some news content on its website for free, but only publishes eight single-topic issues a year, available only on newsstands. This change was the culmination of a long process of de-emphasizing news and current events in favor of focusing on its rankings and research.¹⁵ Rankings, it seems, would no longer "save the franchise." Instead, they have *become* the franchise.

¹² For more on ranking methods see Andrew P. Morriss and William D. Henderson, "Measuring Outcomes: Post-Graduation Measures of Success in the *U.S. News & World Report* Law School Rankings," *Indiana Law Journal* 83 (2008): 791; Michael Sauder and Wendy Espeland, "Strength in Numbers? The Advantages of Multiple Rankings," *Indiana Law Journal* 81 (2006): 205–27.

¹³ For descriptions of the various national and global educational rankings see Alex Usher and Massimo Savino, "A World of Difference: A Global Survey of University League Tables," Canadian Education Report Series, Toronto: Education Policy Institute. See Armin von Bogdandy and Matthias Goldmann, "Taming and Framing Indicators: A Legal Reconstruction of the OECD's Programme for International Student Assessment (PISA)," this volume, pp. 52–85.

¹⁴ See Dora R. Bertram, "Annotated Bibliography: Ranking of Law Schools by *U.S. News & World Report*" (n. 7).

¹⁵ Jeremy W. Peters, "U.S. News & World Report to End Monthly Publication," *New York Times*, November 6, 2010, B3.

Data and methods

Our original research focused primarily on *USN* law school rankings. Our data include approximately 170 in-depth interviews with law school administrators, faculty, students, and leaders of professional organizations devoted to legal education. In addition to interviews, we visited seven law schools varied in rank, region, and mission. We also conducted fieldwork at professional meetings, collected organizational documents, observed online forums devoted to prospective law students, and performed statistical analyses of admissions data.¹⁶

To supplement our findings on law schools, we conducted a smaller comparison study of American business school rankings. This data includes 30 in-depth interviews with upper-level administrators at business schools and organizational documents. Also, relying mainly on a growing body of scholarship as well as collected media reports, we compare our findings about law schools to the effects of rankings of American universities and colleges and world rankings of universities and business schools.

The dynamism of indicators

Here we highlight three general processes by which indicators develop “lives of their own.” First, we discuss how the activity that is fundamental to the construction of indicators—the identification and classification of qualities as commensurate or comparable—necessarily affects how phenomena are represented and cognitively framed. Secondly, we explain how indicators produce “reactivity”: responses on the part of those being measured which result from being observed, compared, and evaluated. Thirdly, we discuss factors that influence the reception of indicators, showing how the meanings of indicators can be recast according to group interests or competition from other evaluations. Each of these processes helps us understand how the effects of indicators diverge from the intentions of those who created them.

Commensuration and cognition

A notable characteristic of indicators is that they commensurate. Commensuration contributes to the dynamism of indicators by changing how people think about things and how information moves around the world. Commensuration is a process fundamental to measurement that entails turning qualities into quantities that share the same metric.¹⁷ Prices, standardized test scores, polls, and votes are common

¹⁶ For more detail on our data and methods, see Espeland and Sauder, “Rankings and Reactivity” (n. 7); Sauder and Espeland, “Discipline of Rankings” (n. 11).

¹⁷ For more on commensuration, see Wendy Nelson Espeland, *The Struggle for Water: Politics, Rationality and Identity in the American Southwest* (Chicago, IL: University of Chicago Press, 1998);

examples of commensuration in which value, ability or knowledge, opinions, and political preferences are turned into numbers. Rankings—whether of students, the risk of doing business in a particular country, the environmental record of corporations, or the quality of universities—are other examples of commensuration.

Commensuration removes information from the particularity of its context and standardizes it. It is a powerful form of simplification that vastly reduces the volume of information that people must attend to, and imposes a common form on, the information that remains. Commensuration simultaneously integrates and distinguishes the entities that it encompasses. The shared metric constructs a common relationship among the entities subsumed under it; at the same time it assigns to each entity a specific, distinctive location that confers a precise relationship on all the other things that it encompasses. Commensuration transforms difference among entities into magnitude on a metric—a matter of more or less rather than of kind.

One reason why commensuration is an important feature of indicators is because it affects our cognition in specific ways. Commensuration makes comparisons easy, almost inevitable. By excluding so much information, it encourages people to scrutinize closely what remains—the numbers—to notice even small numerical differences and to compare them. And by comparing numbers over time, commensuration creates “trends” that can motivate action. Because the difference among entities is turned into intervals on a metric, commensuration creates a clear hierarchy of more or less, better or worse. At the same time that commensuration focuses attention on numbers, it discourages us from noticing differences that are not reflected in the numbers, effectively obscuring those features. We tend not to think about the information that has been excluded, the underlying assumptions that shape what is included, or how much work it takes to produce those numbers.

Rankings transform law schools into numbers, which reinforces the idea that law schools are all the same kind of thing. Rankings emphasize a single similarity (law schools that are accredited by the American Bar Association) rather than their many distinctive qualities (e.g., the types of students they train, the fields of law in which they specialize, whether schools are private or public institutions, or their geographical location). A quick glance reveals the precise relationship of every law school to every other law school; so, for example, in the 2010 rankings list Yale Law School 1st, Brigham Young University Law School 42nd, and the University of New Mexico Law School 67th.¹⁸ It is hard not to see Yale as “better” than BYU, which is “better” than UNM; moreover, the precise difference between BYU and Yale is 41 and between UNM and BYU, 15. Rankings obscure distinctive characteristics of these schools that are not captured by numbers. They do not tell us that Yale trains many future law professors, that BYU incorporates a strong religious and ethical

Wendy Nelson Espeland and Mitchell Stevens, “Commensuration as a Social Process,” *Annual Review of Sociology* 24 (1998): 312–43; Espeland and Sauder, “Rankings and Reactivity” (n. 7) and Wendy Nelson Espeland and Mitchell L. Stevens, “A Sociology of Quantification,” *European Journal of Sociology* 49 (2008): 401–36.

¹⁸ <<http://www.usnews.com/sections/rankings/index.html>>.

mission, or that UNM is committed to training Native Americans and developing the field of Indian Law. The more we focus on rankings, the less likely we are to focus on these or other distinctive aspects of these three fine law schools.

Commensuration produces numbers that move about differently and more easily than other kinds of information. Whereas local knowledge seems cumbersome, contingent, and personal, the stuff of narratives and long descriptions, abstract knowledge like rankings seems portable, impersonal, and easy to transmit. The portability of numbers contributes to their taking on a life of their own. Numbers are de-contextualized abstractions and as such are easy to circulate and insert in new places. Being a “top ten” school is less difficult to advertise, say, than more complex merits. Moreover, we assume that numbers do not require translation, that their interpretation is clear and universal for numerate people. This feature makes them amenable to people who want to use them in new ways. The radical simplification of rankings makes it possible to summarize the entire field of legal education in the United States—or with global educational rankings, the best universities in the world—in a few pages. And it makes it easy for others to email, remember, excerpt, compare, re-organize, or market these numbers.

A second important feature of indicators is that they inevitably contain tacit assumptions about what they are trying to measure and how the world works. As Davis, Kingsbury, and Merry argue, indicators used to evaluate performance embody a “theoretical claim” about what are appropriate standards of evaluation.¹⁹ These implicit claims inform the many processes of commensuration that make up the various factors used to produce ranking; they reflect numerous, often obscure, decisions about just what is being measured, what is included or excluded in a particular indicator, what is desirable, and what it means to be excellent or not.

School rankings disaggregate excellence into discrete components: reputation, selectivity, resources, and placement. They are algorithms that are premised on a particular image of an ideal law school. Schools are measured in terms of how well or poorly they conform to this ideal. For example, merit is defined as high test scores and grade averages. Addressing the role that selectivity plays in undergraduate colleges in the United States, Mitchell Stevens describes the irony of having a school’s stature depend heavily on the students it does not admit.²⁰

Like all indicators, rankings exclude many qualities that educators think are important. The educators we interviewed generally agreed that the factors *USN* incorporates are significant, but they also widely agreed that these are not the *only* characteristics that are important to consider when determining school quality. Administrators listed a wide range of factors that were not included in the *USN* formula, ones that they believed were essential for evaluating schools, such as

¹⁹ Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, “Global Governance by Indicators,” this volume, pp. 3–28.

²⁰ Mitchell Stevens, *Creating a Class: College Admissions and the Education of Elites* (Cambridge: Harvard University Press, 2007).

teaching, faculty scholarship, student and faculty diversity, cost, social climate, and location.²¹

Rankings also imply that law schools share the same set of goals and vision about what “excellence” means. The definition of excellence that informs rankings is based on elite law schools, such as Yale, which reject the vast majority of applicants (last year, Yale turned down 93–94 percent of applicants), only admit students with stellar test scores and grade averages, and enjoy strong national reputations and generous budgets. This conception of excellence does not capture BYU’s ethical goals or UNM’s commitment to training Native American lawyers. Schools that conform to *USN*’s definition of excellence have a big advantage over schools whose objectives depart from it by prioritizing diversity, instruction, or access for disadvantaged students over reputational standing and students’ entrance scores. The dominance of this single, powerful definition of law school quality creates strong incentives for schools to become more like what *USN* measures, which, as many administrators have pointed out, discourages institutional heterogeneity and many forms of innovation. Some schools seem to be abandoning missions that are not rewarded by *USN*’s ranking criteria—most notably, providing opportunities to non-traditional students—because of pressures to rise in the rankings.²²

Another critical assumption embedded within rankings is that school quality is strictly relative, a zero-sum relationship. The rise in rankings of one school requires the decline in rankings of others. Depending on a school’s rank and the magnitude of the shift, one school’s change in its ranking could affect the rankings of many other schools. Unlike categorical evaluations, where multiple schools could be evaluated as excellent or good, accredited or not, the linking of schools’ fates in this way encourages schools to scrutinize not only their numbers but also the numbers of many other schools. A school’s relative “performance” could decline, even if things stayed the same or if a school improved on some measures but not as much as other schools. This kind of scrutiny has led to what some administrators characterize as an “arms race” in rankings, where the pressure to focus on rankings factors and devise creative strategies for boosting them becomes more intense over time.

Along with core assumptions about measurement or excellence, rankings also reflect many relatively trivial decisions. For example, 12 percent of a school’s ranking is determined by how many of its graduates are employed nine months after graduation. Schools often send repeated requests to graduates to get this information and must report to *USN* how many students fail to respond. *USN* assumes that two-thirds of non-responding students are unemployed. Because scores are so tightly bunched together, this assumption, however reasonable, can have a dramatic impact on a school’s ranking. As one dean of career services told us:

²¹ These factors appear in a letter sent by the Law School Admissions Council to every student who registered to take the LSAT for ten years. It was signed by 174 of 192 deans of accredited law schools in 2007 and used to appear on the LSAC website. At the time of writing, it no longer does.

²² Wendy Espeland and Michael Sauder, “Rankings and Diversity,” *The Southern California Review of Law and Social Justice* 18 (2009): 587–610.

We know that if one of your students doesn't report one year, your employment ranking will go down, not because your student's unemployed but because the students wouldn't report. And that will have an impact on how the school's ranked, and you will have to explain it to the alums when they just look at the ranking. [Small changes can] have an enormous impact because everything's at the margins. So many schools are tied in these rankings that you make a marginal switch here or a marginal switch there, you can certainly fall [a long way] in these rankings.

Rankings encourage people to think that the most salient differences among schools track a standardized definition of quality that pits schools against each other.

A third feature of indicators closely related to commensuration is that creators of indicators usually favor things that are easier to measure or things for which measures already exist. In constructing an indicator, it is less complicated to include median test scores, the grade averages of entering students, or the faculty-student ratio than it is to devise measures for the quality of instruction, faculty scholarship, diversity, or school culture. And things that are relatively easy to measure are usually things that have already been standardized in some way. Qualities that are hard or contentious to measure, or qualities that do not already have rules or procedures that help to standardize them, are more likely to be excluded from indicators. For example, nearly everyone would agree that a good library is an important feature of an excellent law school. The characteristics of a good library might include an appealing building, plenty of study space, top-notch research librarians, and sophisticated electronic resources, as well as a large and systematic collection. But in measuring library quality as part of its "faculty resources" factor, *USN* rankings include only the number of volumes in the library because this is the most readily quantifiable aspect of library quality.

The processes we have described above help us understand the dynamism of indicators by showing how these properties change what we pay attention to, what we ignore, and the form and extent to which information circulates. These features are shared, at least to a degree, by all indicators. But it is also important to consider how distinctive properties of an indicator contribute to its trajectory. For example, the power of *USN* rankings is partly a function of their origins in a media company with long experience and many outlets for distributing rankings. This expertise permitted *USN* to publicize and disseminate rankings to a wide audience. Furthermore, that they were produced by a news magazine helped to frame annual rankings as "news," which fostered interest in them from other media. In investigating indicators, it is important to account for what they share with other types of indicators as well as features that reflect their unique contexts.

The reactivity of indicators

A second important general process that influences the dynamism of indicators occurs through the reactions that they evoke from the people who use or confront them. Indicators produce "reactivity" because social measures prompt individuals and the organizations they inhabit to change their behavior in response to being

measured or evaluated. A survey question may stimulate us to consider an issue we had previously ignored, or the task of counting something might prompt us to keep better records at work. Sometimes, changing behavior and thinking is the purpose of an indicator. Performance measures, whether created by firms assessing their employees or NGOs evaluating the environmental records of firms, are designed to improve performance as well as evaluate it. We want employees to set and meet goals, and we want firms to focus on their environmental impacts. But changes in behavior can also be independent of—or in the language of organizational theory, “decoupled from”—the processes the indicators are designed to reflect or measure. When this happens, people focus on the indicator rather than the underlying behavior or quality the indicator is designed to assess. This is a common pattern in people’s reactions to indicators and an important reason why indicators come to seem autonomous.

Because rankings seem like such clear and authoritative indicators of performance, and because they are publicized so broadly, law schools face intense pressure to improve the measures that comprise rankings. Deans often report facing dilemmas about choosing between doing what is best for improving the indicator or doing what they deem best for the education of the students. Many believe that efforts to improve rankings have come at the expense of improving legal education. As one law school dean explained:

[Law schools] feel pressured into making decisions that from an educational point of view they would really prefer not to make. They are not doing what their professional judgment tells them is good for their students. They’re doing what their statistical judgment tells them is good for their rankings.

Many law schools, for example, have dramatically increased spending on marketing in an attempt to boost the reputational factor that determines 40 percent of their ranking. This marketing, often costing hundreds of thousands of dollars, entails mailing eye-catching brochures touting school and faculty achievements to peers and practitioners who might fill out the following year’s reputational survey for *USN*. Colin Diver, president of Reed College and the former dean of the University of Pennsylvania Law School, described this practice:²³

When I was law school dean, my mailbox would begin to fill up about a month before the *U.S. News*’s annual ‘beauty contest’ questionnaire arrived—with glossy admissions brochures, alumni magazines, lists of faculty publications, and breathless announcements of new buildings and academic symposia, all accompanied by bland cover letters from my counterparts expressing the thought that I might find the enclosures interesting and illuminating. In my ten years as dean I only once received a cover letter that came right out and said what every other letter wanted to say: ‘When *U.S. News* opinion survey comes out next week, please keep our law school in mind.’

While most administrators we interviewed noted the ineffectiveness of this strategy—explaining that they simply throw away this material away without reading

²³ Colin Diver, “Is There Life After Rankings?”, *The Atlantic* (November 2005), 136–9.

it—they also feel compelled to follow suit so as not to appear that they are doing less than their peers to address the school's rank. And money spent on marketing is unavailable for other uses.

Marketing aimed at boosting rankings is just one common example of how resources are deployed differently because of concerns about rankings. An even clearer example of behavior being driven by indicators is when people attempt to “game” the indicator. Gaming strategies reflect actors' blatant efforts to improve their standing on an indicator in ways that are clearly unconnected to—or at times even flaunt—the designated purpose of the indicator. Gaming strategies are aimed almost exclusively at manipulating accounting practices and procedures, changing numbers without changing anything else. Some law schools were caught misreporting their median LSAT scores to *USN*. Other schools changed their definition of what counts as employed—from those employed in the legal profession to those with any job in any sector (taxi driver, nanny, part-time intern in the law library)—to raise their placement statistics. Schools that rejected this strategy lost ground in the rankings to those who used it, so now nearly every school has adopted this inflated definition of employment. Still other schools have admitted low-scoring students to their night or part-time programs (when *USN* did not require them to include these students in their reporting) for a short period before transferring them to the full-time program, a strategy that allows these schools to maintain their revenue and possibly their diversity profile without losing ground on their median LSAT scores or GPAs. Some schools hire their unemployed graduates as research assistants until after they report their numbers. To varying degrees, these practices show how the pressure to improve rankings encourages actors to interpret and act on rankings in ways that reinforce the distance between the factors that make up the indicator and the qualities it measures. This tendency to treat abstract ideas as if they were powerful, autonomous things is more broadly known as reification.

The mediation of the reception of indicators

A third factor to consider when analyzing the dynamism of indicators concerns the structures and processes that mediate their reception by external audiences. One important feature of reception is to appreciate that indicators have multiple audiences or constituents. This matters because the features of a particular audience and how they use numbers shape the meaning of those numbers. We like to think that numbers have stable, transparent meanings but this conception does not square with practice; groups notice and interpret indicators differently depending on their experience, interests, and their location within organizations and societies. Consequently, the same indicator can be understood and used in a variety of ways by different audiences and this variety shapes its authority and trajectory.

Most law school administrators and faculty believe that rankings can be arbitrary, that many of the distinctions embodied and reproduced in rankings are a product of methodological choices rather than a real reflection of differences in quality. Most prospective students, on the other hand, are far less critical of the numbers,

believing that they represent real differences among the schools. At least some see rankings as if, as one dean of admissions put it, they were “written in stone,” and will choose one school over another because it is ranked a few spots higher that particular year. A current law school student explained,

I know I wasn't concerned with methods. I just figured [*USN*] knew what they were talking about, and I didn't know any way to compare schools and so I wasn't concerned about measuring errors or how accurate the rankings were. I knew that this is what everybody was looking at and what everybody else is using, so I was willing to take [*USN*'s] word for it.

Still other audiences such as university administrators, trustees, and employers may publicly acknowledge the shortcomings of the rankings but nevertheless use them in benchmarking, making comparisons among divisions or departments, for allocating resources, or as a way to create thresholds to screen job candidates.

Sometimes divergent interpretations of rankings contradict one another, which can lead to conflict over which views will prevail in discourse and practical use. Such conflict occurred with respect to law school rankings. Initially, administrators adamantly resisted the rankings, considering them too flawed to be of value; however, because they were useful to influential external constituents (especially prospective students, employers, and university administrators), law school administrators were soon forced to adapt to the views of these other constituents for fear of losing their jobs or doing damage to their schools' reputations.²⁴ The moral here is that once an indicator is released into the world, it can be interpreted and used in a number of ways and these different interpretations can influence one another. Despite its seemingly objective nature, the meaning of a quantitative indicator is still negotiated by the audiences that use it rather than contained firmly within itself. This potential for varying interpretations highlights the idea that indicators sometimes escape from the realms of their intended purpose and are adapted by groups to serve their own needs. Organizational scholars would characterize this pattern as indicators becoming “solutions” in search of a problem instead of the other way around.²⁵

A second factor that influences the reception of indicators is the degree of competition among indicators within a field. When a single indicator dominates a particular field, as *USN* rankings dominate the field of legal education, this indicator wields great influence, is more likely to be taken at face value by outside audiences, and will often elicit more and stronger reactions as efforts are concentrated on improving a single number. Multiple indicators in a field produce different effects. Our comparative work on business school rankings provides a clear example. In contrast to law school rankings, there are five or six influential business school rankings and a host of others that receive some attention. No single ranking dominates the definition of business school reputation. Further, because all

²⁴ For a more detailed description of this process, see Michael Sauder, “Interlopers and Field Change: The Entry of *U.S. News* into the Field of Legal Education,” *Administrative Science Quarterly* 53 (2008): 209–34.

²⁵ Michael Cohen, James March, and Johan Olsen, “A Garbage Can Model of Organizational Choice,” *Administrative Science Quarterly* 17 (1972): 1–25.

of the rankers use different formulae to determine the relative standing of business schools, the rankings that they produce differ markedly from each other—some (like *Businessweek*) focus on recruiter and student satisfaction, some (like *USN*) on reputation ratings and quantitative measures, some (like *The Wall Street Journal*) on employer surveys, and some (*Financial Times*) on graduate salaries.

While competing rankings have not made business schools immune to the effects of rankings—they still prompt business schools to redistribute resources, change organizational activity, and devise gaming strategies—they have minimized the impact of certain key problems the rankings create by producing a more ambiguous signal about the precise standing of these schools. This occurs in some predictable ways, such as weakening the effects of small differences between schools and small changes in rank. While many law school deans worry that random statistical variations in *USN* will hurt their school, business school deans tend to look at how well their schools measure up across a range of rankings and are less concerned about the changes in any one specific ranking. This makes them less vulnerable to the vagaries of any one evaluative system.

The multiple rankings of business schools also provide them with greater reputational control and flexibility. Not only do schools have more opportunities to rank well because there are different rankings with different methodologies, but they can also play to their strengths. An obvious marketing strategy for schools is to advertise the rankings that present them in the best light. Along the same lines, many administrators believe that the ambiguity created by multiple rankings provides business schools with many options for changing or building their future reputations. For example, a school could choose one ranking which best suits their school in terms of its strengths and its goals, and then this school could redistribute resources in order to maximize their standing according to that ranking.

The example of business schools also suggests a less obvious effect of multiple indicators: multiple rankings may encourage consumers of the rankings to be more critical of the ranking process itself. Business schools are evaluated by a multitude of rankers, all claiming to measure the same basic concept—the quality of business schools—but produce much different results in their measurements. This situation permits schools to question the validity of the rankings when dealing with external audiences. As one dean explained:

We're very pleased about every new ranking that comes out that kind of turns the world on its head. *Wall Street Journal* came out with rankings two years ago, and whenever I see a ranking that has Stanford in the 45th spot, [X] spots behind my school, I say that this is good news because it deflates the bubble. It ought to make people think just what the heck this is and what it is that they are really trusting. And whether they should trust the *Wall Street Journal*, which has them 45th or this other ranking that has them 2nd or whether it has them 18th, it ought to open up the eyes of most people who should say, "Ah, I get it now. I could do my own ranking tomorrow and make it come out any way I want, depending on what I think is the criteria."

In short, these competing evaluations undermine the taken-for-granted validity of the rankings that exists in the field of legal education. In doing so, the multiple

rankings also promote the idea that there are many ways to create a quality business school. These examples of factors that mediate the reception of indicators further illustrate how the effects generated by indicators cannot be controlled by those who create them; quantification triggers processes that are hard to predict and can have unanticipated consequences.

Global education rankings

We have discussed three general processes that play an important role in determining the type of effects indicators generate and how they correspond or depart from the purposes for which they were designed. Although our examples to this point have been limited to law school rankings in the United States, we believe that these patterns apply to social indicators more generally. We illustrate their extension to global indicators by showing their relevance to the burgeoning field of worldwide education rankings.

There are dozens of prominent national rankings, also called league tables, throughout the world, and a growing number of global university rankings that compare higher education institutions.²⁶ Recently, there have been two clear leaders in this field: the QS World University Rankings (QSWUR)²⁷ and the Shanghai Jiao Tong Academic Ranking of World Universities (ARWU).²⁸

Briefly, the QSWUR has produced its ranking of “the top 200 world universities” since 2004, along with a series of more specific field rankings.²⁹ Claiming to survey over 2,500 institutions, QSWUR places a heavy emphasis on assessments of reputation by other academics and employers/recruiters; these measures make up 40 percent and 10 percent of a school’s overall score, respectively. Student–faculty ratio (20 percent), citations per faculty (20 percent), and international representation among students and faculty (10 percent) constitute the remaining half of

²⁶ For overviews of different world rankings, see Alex Usher and Massimo Savino, “A World of Difference: A Global Survey of University League Tables,” in *Canadian Education Report Series* (Toronto: Education Policy Institute, 2006); Jamil Salmi and Alenoush Saroyan, “League Tables as Policy Instruments: Uses and Misuses,” *Higher Education Management and Policy* 19 (2007): 31–68.

²⁷ From 2004 to 2009 the QSWUR rankings were published jointly by Quacquarelli Symonds Limited and the *Times Higher Education Supplement* (THES). Beginning in 2010, QS and THES ended their partnership, with QS continuing to publish the same world rankings using the same methods, and THES developing a new system of world rankings in conjunction with Thompson Reuters. These new rankings, which we label THES rankings, were also first published in 2010. USN began publishing the QS rankings of the “World’s Best Universities” in the United States in 2010. Our discussion refers to the pre-2010 rankings unless otherwise noted; here QS refers to its 2010 rankings and THES refers to its revamped 2010 rankings.

²⁸ The ARWU rankings are widely considered the “brand leader.” See Ellen Hazelkorn, “Learning to Live with League Tables and Ranking: The Experiences of Institutional Leaders,” *Higher Education Policy* 21 (2008): 193–215. *Asiaweek* published university rankings of Asia-Pacific schools for four years until it folded in 2001.

²⁹ <<http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/institution-inclusion>>.

the ranking.³⁰ The ARWU rankings, published by Shanghai Jiao Tong University, began in 2003. Ranking the top 500 universities, the ARWU constructs its indicator from a set of “objective measures” of research productivity.³¹ These include current faculty and alumni winners of Nobel Prizes and Field Medals (respectively, 20 percent and 10 percent, of the overall score), citation counts (20 percent), papers published in *Nature* and *Science* (20 percent), papers included in prominent citation indices (20 percent), and per capita academic performance (10 percent).³²

Commensuration and global rankings

Commensuration plays a central role in the dynamism of global educational rankings, just as it does for law school rankings. Like other rankings, the QSWUR and AWRU rankings turn universities into numbers via processes of commensuration; historically, culturally, and geographically disparate educational institutions are made comparable by virtue of their shared metric. To the extent that the numbers come to define comparisons among institutions and assessments of quality, it is irrelevant whether a university is 500 or 50 years old, is located in Tokyo or Tucson, was founded by the state or subsidized by an oil tycoon.

As decontextualized numbers, rankings circulate quickly and broadly. Their annual release is the subject of extensive media coverage in newspapers, magazines, and on Internet blogs, and the competition among various rankings has become fierce as their number and influence proliferate. Newspapers routinely report how national or local universities fare in the global ranking, and rankers continue to innovate in the dissemination of their products. The QSWUR, for example, which claims to be the “most popular rankings in the world” with 20 million annual visits to its website, has recently released an iPhone App that allows users to view its rankings in multiple forms and to discuss rankings with other users.³³ QS reports that each year about 1,000 different newspapers, journals, and websites refer to their rankings.³⁴ There are now numerous blogs dedicated solely to the discussion of rankings, including those sponsored by rankers (e.g., QS and *USN*) and those monitoring the quality of rankings (e.g., World University Watch).³⁵ Regardless of the particular media or forum used, it is clear that annual world rankings are “news”

³⁰ <<http://www.topuniversities.com/world-university-rankings/understanding-qs-world-university-rankings-methodology>>.

³¹ <http://www.arwu.org/ARWU_2010_press_release.jsp>.

³² <<http://www.arwu.org/ARWUMethodology2010.jsp>>.

³³ <<http://www.topuniversities.com>>.

³⁴ <<http://www.topuniversities.com/world-university-rankings/understanding-qs-world-university-rankings-methodology>>.

³⁵ Cites include: <<http://www.topuniversities.com/university-rankings/world-university-rankings/qs-world-university-rankings-blog>>; <<http://www.usnews.com/blogs/college-rankings-blog/index.html>>; <<http://rankingwatch.blogspot.com/>>.

that spreads quickly and widely, and is scrutinized closely in ways that are made possible by commensuration.

Just as law school rankings reflect implicit “theories” of educational excellence, so too do global rankings. Excellence is understood as relative, as something that can be universally defined using some weighted total of discreet components, components which in turn reflect the assumptions of those producing them. That the same standardized notion of educational quality should be appropriate for universities spanning six continents with profoundly different populations, trajectories, and goals is, of course, a radical presumption, even more radical than assuming that all American law schools are comparable.³⁶ In their survey of approximately thirty of the most prominent national and international rankings (including the QSWU and AWRU rankings), for example, Usher and Savino conclude, “Clearly, the choice of indicators and the weight given to each indicator make an enormous amount of difference to the final output. Indeed, it is no exaggeration to say that when publishers advertise their product as a guide to ‘the best’ institutions, it is the publishers themselves who largely decide the best simply through their choice of indicators and weightings.”³⁷

Critics of global rankings have noted: the QSWUR’s and AWRU’s nearly single-minded focus on research measures;³⁸ their relative neglect of undergraduate education—they include no direct measures of teaching or student educational gains or “outcomes”;³⁹ their biases in favor of English-speaking countries as well as old, wealthy universities with stable English names;⁴⁰ and the American model of university education.⁴¹ The AWRU, for example, clearly favors schools with strong science and technology programs.⁴² Others suggest that important educational values are ignored by the conception of quality embedded in these rankings. For example, an emphasis on egalitarianism or a commitment to educating a broad public helps explain why countries such as Germany and France fare relatively less well in these rankings.⁴³ And it is certainly easier to count recipients of Nobel

³⁶ As Billaut, Bouyssou, and Vincke (2009: 27–8) note in their critique of the AWRU rankings, among the assumptions embedded in rankings is the idea of what a university is, something more complex in Europe than in the United States or Britain. France, for example, includes public universities, which vary by the programs available and the types of degrees granted, the *Grandes Ecoles*, which mostly grant masters degrees, and prominent research centers shared by various institutions. The 2003 AWRU rankings included the “College de France” which has no students and does not grant degrees. “Should you Believe in the Shanghai Ranking? An MCDM View,” *Scientometrics* 84 (2003): 237–63.

³⁷ Usher and Savino, “A World of Difference: A Global Survey of University League Tables” (n. 26).

³⁸ See David D. Dill and Maarja Soo, “Academic Quality, League Tables and Public Policy: A Cross-National Analysis of University Ranking Systems,” *Higher Education* 49 (2005): 495–533. The new THES rankings include a teaching component.

³⁹ Salmi and Saroyan, “League Tables as Policy Instruments” (n. 26), 31–68.

⁴⁰ Billaut et al., AWRU Ranking (n. 36).

⁴¹ Salmi and Saroyan, “League Tables as Policy Instruments” (n. 26), 31–68.

⁴² See, e.g. “World University Rankings,” World Education News and Reviews, Vol 10:4, accessible at <<http://www.wes.org/ewent/PF/06aug/pfeature.htm>>.

⁴³ See Ellen Hazelkorn, “Rankings and the Battle for World-Class Excellence: Institutional Strategies and Policy Choices,” *Higher Education Management and Policy* 21 (2009): 1–22.

Awards and Fields Medals for mathematics than it is to measure quality directly or expansively.

As quantitative, hierarchical, and heavily edited abstractions of quality, the features of global educational indicators, like their national law school counterparts, help us understand their dynamism. Global educational rankings focus attention in particular ways, travel widely, and reinforce some conceptions and values at the expense of others. In doing so, they help to remake images of the purpose of education and help to redefine institutional status in ways that favor some but affect all.

The reactivity of global educational rankings

Global educational rankings elicit a range of reactions from different groups, reactions that help propel them to new uses and new meanings. The visibility of global rankings and the precise distinctions they create encourage a careful parsing of their components. Intense pressure to improve a school's position in the rankings means that efforts often are directed toward improving the indicators rather than the qualities they are intended to measure. Each of these processes, but especially the decoupling of indicators from their intended purpose, reinforces the authority and autonomy that people attribute to rankings.

In a study based on surveys of administrators of higher education institutions around the world, Hazelkorn shows the wide range of universities' reactions to global rankings.⁴⁴ Administrators and faculty face strong pressure to improve "numerical results"—to emphasize work that is linked to the components of rankings. Some universities have created new administrative positions dedicated exclusively to improving the school's rank⁴⁵ while others have increased staffing in international offices in an effort to improve the representation of international students (an important criteria in the *Times Higher Educational Supplement Rankings*).⁴⁶ Universities have also devised strategic plans with the explicit goal of improving a school's rank. For example, Professor Du Chao-Hui, the head of graduate education at Shanghai Jiao Tong University, the birthplace of the AWRU rankings, recently noted that as part its 2020 strategic plan, his university's goal was to become a top 100 world university and that the university will use rankings to "guide our own goals and target our ambitions correctly." Whatever the University is doing seems to be working. The school was in the top 500 in 2003 (401st) but had become a top 300 school by 2010 (201–300). Du also noted that two Chinese universities—Beijing and Tsinghua—have moved into the Top 200, and that

⁴⁴ See Ellen Hazelkorn, "Impact and Influence of League Tables and Ranking Systems on Higher Education Decision Making," *Higher Education Management and Policy* 19 (2007): 87–110.

⁴⁵ Hazelkorn, "Learning to Live with League Tables" (n. 28).

⁴⁶ Institute for Higher Education Policy [IHEP], May 2009, "Impact of College Rankings on Institutional Decision-Making: Four Country Case Studies," <<http://www.ihep.org>>.

China now has 22 universities in the top 500, an increase from eight in 2005.⁴⁷ In Malaysia, after two universities dropped in the QSWUR rankings, the Prime Minister created a national committee to evaluate how the country could raise the stature of its public universities.⁴⁸

Hazelkorn offers stark examples of how resources are being redistributed in reaction to rankings.⁴⁹ Some schools are diverting funding to areas that are more likely to produce higher rankings according to the criteria used by the rankers. Because research in particular areas—most notably, the natural sciences—is best captured in the international databases used to measure publication and citation success, areas not represented as well in these databases (e.g., the arts, humanities, and social sciences) or areas without the same peer-reviewed article traditions (education, business, and engineering) have experienced reductions or eliminations in funding because they do not “count in the rankings.” Germany recently announced that it would allocate 2 billion Euros among its universities, based mainly on their research strength.⁵⁰

Other schools are undergoing even more extensive restructuring in efforts to increase their visibility and reputation scores in the rankings. To compete with the large US universities, these schools concentrate resources by either merging departments and institutions to increase their size and visibility or separating undergraduate from graduate education in order to develop very visible research institutions.⁵¹ These changes affect the training provided to undergraduates and illustrate a general emphasis on research at the expense of instruction.

Another form of redistribution takes place at national levels. Some countries, such as Australia, have chosen to allocate more funds to a few select universities that have a chance to reach the top 100 at the expense of funding more universities across the country.⁵² Japan is considering changing the role of smaller regional universities as support streams to larger universities either by closing them or turning them into second-class “teaching universities” distinct from larger, more prestigious research-driven universities.⁵³ Critics fear that the spread of policies like these will foster the development of a caste-like division of universities within many countries. By encouraging university systems to resemble “winner-take-all” markets where countries pour resources into a few select universities instead of distributing them to benefit their populations more broadly, rankings threaten long-held values about the role of public education.⁵⁴

⁴⁷ Guy Healy, “Ranking Shock has No Impact,” *The Australian*, September 15, 2010, accessible at <<http://theaustralian.com.au/higher-education/>>.

⁴⁸ Aisha Labi, “Obsession with Rankings Goes Global,” *Chronicle of Higher Education* 55 (2008): 27.

⁴⁹ Hazelkorn, “Battle for World-Class Excellence” (n. 43), 8–17.

⁵⁰ Labi, “Obsession with Rankings Goes Global” (n. 48), 42.

⁵¹ Ibid.

⁵² Ibid., 11.

⁵³ Hazelkorn, “Battle for World-Class Excellence” (n. 43), 13.

⁵⁴ In her research on the effects of global rankings of business schools, see Linda Wedlin, *Ranking Business Schools: Forming Fields, Identities and Boundaries in International Management Education*. (Cheltenham, UK: Edward Elgar, 2006).

These are just a few of the strategies adopted by universities and their countries in response to the world rankings. Some other tactics to boost rankings seem even more decoupled from substantive educational goals. For example, investing in Nobel Prize winners is an expensive proposition and the award often acknowledges work done many years before. Yet some Irish and British schools are trying to increase the number of Nobel recipients on faculty.⁵⁵ Other schools are encouraging faculty to only publish in English-language journals, buying advertising space in *Nature* and *Science*, and reducing faculty size and composition so that a disproportionate number of faculty members are late in their careers and very highly cited.⁵⁶ The educational value of these decisions, of course, can be debated; less debatable, however, is the fact that the rankings are playing an important part in defining what constitutes educational value and motivating schools to change their behavior accordingly.

Mediators in the reception of global indicators

World rankings, despite the seemingly objective nature of the numbers they produce, are interpreted and used differently by different constituents. Hazelkorn's survey of administrators, for example, shows that there are about as many different views of the global rankings as there are stakeholders, which include benefactors, current and future faculty, employers, government officials, and students.⁵⁷ Much like in the United States, those inside the institutions being ranked are likely to be more skeptical of these blunt assessments than are external constituents, who are more likely to take them at face value. But there are also key audience dynamics central to global rankings that are less important in domestic rankings. For instance, international students are much more likely to use the rankings to make decisions about school choice than are students in their own countries.⁵⁸ Administrators pay special attention to this category of students not only because they are key to claims schools can make about being an international university, but also because the proportion of international students attending a university is one measure used by QSWUR rankings to determine a school's overall rank.

Global rankings also bring into play a stakeholder that is virtually nonexistent in the US rankings: national governments. World education rankings are becoming an important indicator for some governments because national identities and reputation are tied to education systems⁵⁹ and because rankings can serve as symbols of world standing or economic strength.⁶⁰ Many governments use international rankings as political tools to lobby for and expedite desired improvements.

⁵⁵ IHEP, "Impact of College Rankings," 6, citing Burton Bollag, "College Rankings Catch on Overseas," *Chronicle of Higher Education* 53 (2007): 38.

⁵⁶ Hazelkorn, "Battle for World-Class Excellence" (n. 43), 16.

⁵⁷ Hazelkorn, "Learning to Live with League Tables" (n. 28), 197.

⁵⁸ Hazelkorn, "Battle for World-Class Excellence" (n. 43), 37.

⁵⁹ Salmi and Saroyan, "League Tables as Policy Instruments" (n. 26), 31–68.

⁶⁰ Hazelkorn, "Battle for World-Class Excellence" (n. 43), 37.

Germany, France, Taiwan, South Korea, India, Indonesia, Denmark, and Japan have all employed international rankings to justify major reforms in higher education while explicitly tying these reforms to the indicators that are most heavily weighted in these rankings.⁶¹ Moreover, some countries rely on the distinctions made by rankings to define other policies as well. Hazelkorn notes that Dutch immigration law favors graduates from “top 150 universities,” Mongolia, Qatar, and Kazakhstan only provide scholarships to students who are accepted at “top 100 institutions,” and Singapore’s Foreign Specialist Institute restricts collaboration with its local schools to “top 100 universities.”⁶² These examples demonstrate both that governments care a great deal about the rankings of their national universities and that their interpretation of these numbers represents a powerful voice that is distinct from global rankings. We can also clearly see how the indicators can influence, rather than merely reflect, what constitutes educational quality as states change their educational systems to conform to ranking criteria and use the thresholds created by rankings to make policy distinctions.

The competing definitions of educational quality embodied in global rankings also influence their reception. The existence of two prominent world rankings should produce some of the same dynamics we discussed above in relation to business school rankings in the United States. Despite the fact that both the QSWUR and AWRU place a similarly strong emphasis on research, the results of these two ranking systems differ significantly: less than half of the schools listed in the top 50 by one ranker are in the top 50 of the other, and less than a quarter are within five positions of each other on both rankings.⁶³ As with business schools, we would predict that this would allow schools concerned about their standing to orient their strategies toward the ranking that best suits their strengths or is most practicable (e.g., hiring Nobel Prize winners rather than beginning a large-scale advertising campaign to improve their peer reputation scores), to deflect criticisms about small changes in rankings, and to highlight the subjective nature of ranking outcomes.

Research by Linda Wedlin on the global rankings of business schools illuminates the role that multiple rankings can play.⁶⁴ She found that leading European business schools lobbied *The Financial Times* to create international rankings of MBA programs in 1999 as a way to counter the influence of media rankings of US business schools and, more fundamentally, challenge the dominance of American models of business education. By including factors that reflected the strengths of European schools, these rankings eventually helped to create the field of “European management education.”

The picture painted by global rankings is also complicated by the existence of a large population of national rankings or league tables. Many countries now have

⁶¹ Ellen Hazelkorn, “Handle with Care,” *Times Higher Education*, July 8, 2010 accessible at <<http://www.timeshighereducation.co.uk/story.asp?storycode=412342&encCode=4772581801BC26960875JTBS737226611>>.

⁶² Ibid.

⁶³ Salmi and Saroyan, “League Tables as Policy Instruments” (n. 26), 47.

⁶⁴ Wedlin, *Ranking Business School* (n. 54).

national rankings⁶⁵ and these must be attended to along with the QSWUR and AWRU. Not only do these rankings have different scopes (national vs global), but it is also common for them to have different points of emphasis as well. In general, national rankings are more oriented toward domestic students and, because of this, tend to measure instructional quality and outcomes rather than focusing solely on research productivity. These differences between the national and global rankings create dilemmas for schools, forcing them to make decisions about the audiences to which they would most like to appeal or how to strike a balance between maintaining educational quality for their local constituents and developing a global reputation that will appeal to their international constituents.

Conclusion

Just seven years old, global educational rankings exert remarkable influence. This influence both reflects and propels forward the internationalization of economies and curricula, and the circulation of students across borders of all sorts. It is buttressed by a deep faith in the ability of markets to evaluate and distribute goods, and in a vision of education as an investment with variable economic and status returns. The dynamism of rankings reflects the dynamism of the commodity form, a form that was borrowed from the thirties, adopted in the eighties, and adapted and spread in the nineties as both private and public performance indicators.

What began as a way for an American magazine to remake its tired image has morphed into a network of organizations and a set of practices that have, quite literally, changed how the world thinks about higher education. As authoritative measures of educational status and quality, rankings have accumulated a range of constituents—from parents, to firms, to sovereign governments, to the European Union, which is currently investigating producing its own educational ranking. Many in this wide swath of constituents embrace the shortcuts rankings offer to make judgments about quality and merit. As people in new places adopt new uses for rankings, create offices and policies to manage them, devise new rankings to counteract old ones, and come to see all universities in relation to each other, the authority of rankings—for better or worse—becomes ever more secure. If particular rankings are suspect, rankings as a trope of evaluation now seem inevitable, a natural feature of the landscape. Few can escape their influence.

Especially at the international level, the influence of these powerful definitions of educational quality raises concerns about the effects of applying a single standard to such radically diverse institutions. Not only do these universities have different histories, goals, and missions,⁶⁶ but they also vary greatly according to such

⁶⁵ For a comparison of various countries' national rankings and league tables, see Dill and Soo, "Academic Quality" (n. 38).

⁶⁶ Marijk Van der Wende, "Quality Assurance in Higher Education and the Link to Internationalisation," available at <<http://www.ipv.pt/millennium/wende11.htm>>.

characteristics as “sources of funding, government patterns, degree of management autonomy, differences in institutional missions, availability of reliable data and potential for manipulating statistics.”⁶⁷

Some researchers and policy experts believe that rankings have positive as well as negative consequences. They can be used for setting benchmarks, increasing transparency, allocating resources more efficiently, stimulating discussions about how to measure institutional success, or encouraging countries to boost their aspirations.⁶⁸ Our response to advocates is that more attention should be paid to the unanticipated or latent consequences of these global rankings, and how these impacts vary in different locations, as well as to research that carefully looks inside organizations to track what people do as well as what they say. Important questions for scholars include: “Are the rankings making universities more homogenous?,” “Which constituents benefit from this process and which are harmed?,” “How is the content of education changed?,” “Which types of universities are most affected?” These are questions that motivate much of the analysis in this volume, analysis that invites corroboration, specification, and debate.

We have argued that the processes and characteristics that help propel rankings are shared by many indicators. Indicators commensurate, changing what we notice and how we think about something by highlighting, excluding, and flattening difference. People at times react to indicators in ways that recreate the objects they measure, amplify their influence, or undermine their purpose. In addition, these reactions change over time. Audiences respond differently to indicators depending on the nature of their investments, their historical contexts, and whether there are single or multiple indicators in a given domain. Some groups may ignore them, some may obsess over them, and some may play various indicators off against one another. Attending to these features of indicators and the various responses they evoke in their audiences will help scholars understand the implications of this growing but complex form of governance.

Going forward, a comparative approach will be necessary to develop a thorough understanding of the effects and trajectories of indicators, to disclose which features of indicators are distinctive and which are shared, or to explain how these variations are linked to particular contexts. Comparing the sources of global rankings—commercial media, national governments, or international organizations—will provide insight into how politics affect circulation, visibility, and legitimacy. Comparing the effects of indicators over time will help reveal processes of institutionalization and how effects change as they become regarded as inevitable. Comparing how various audiences interpret and come to understand these evaluations will illuminate the processes by which meanings are negotiated in different contexts and how this shapes durability and credibility. Comparing successful indicators with failed ones might reveal the links between different bases of authority and an indicator’s success. And perhaps most importantly, comparing indicators within

⁶⁷ Salmi and Saroyan, “League Tables as Policy Instruments” (n. 26), 58.

⁶⁸ Hazelkorn, “Learning to Live with League Tables” (n. 28), 200; IHEP “Impact of College Rankings” (n. 55), 1.

and across fields will help us understand who are the winners and losers with a given indicator, whose expertise increases and whose declines, what indicators highlight and what they obscure. While all successful indicators are dynamic, the particular form that this dynamism takes depends on these relationships and processes. As our work makes clear, governance by indicators, like all forms of power, is a moving target.

Semiotics of Indicators: The Case of Corporate Human Rights Responsibility*

Ronen Shamir and Dana Weiss

Introduction

The global human rights regime aims to provide defenses and remedies to victims, prevent violations by potential perpetrators, and promote the protection of human rights. Since the mid-1990s, this global regime has expanded to include the role and duties of multinational corporations, framed in terms of human rights violations that occur as a direct or indirect result of corporate activity. The principled treatment of this problem shifts attention from the conventional focus on the actions and policies of state and quasi-state apparatuses towards greater concern with the human rights responsibilities of corporations.

This chapter examines some aspects of the drive to enhance the responsibilities of corporations. We do not claim exclusivity and comprehensiveness. There is more than one way through which to explore this drive. For example, we are not dealing here with the legal dimensions of “corporate human rights responsibility.”¹ Rather, we focus on aspects of the so-called “soft law” solutions to the problem: self-made and self-imposed corporate policies and practices, regulation by private authorities, and theories, technologies, and instruments of business management. By and large, we posit the existence of “corporate human rights responsibility” as a set of ideas and practices that aims to diagnose and treat corporate-related human rights violations (actual or potential). This chapter is mostly descriptive. We explore the mechanics of assembling a framework that promotes the human rights responsibilities of corporations. We trace the ways and means through which this framework is deployed and sustained, and assume the involvement of various institutions, initiatives, instruments, symbols, rituals, procedures, and texts. We assume that all these artifacts and entities are at least potentially mobile, circulating around, making connections, and enabling all kinds of social forms and assemblies.

* Prepared for the “Indicators as a Technology of Global Governance” Conference, NYU, September 13–14, 2010 (revised for publication June 2011).

¹ Steven R. Ratner, “Corporations and Human Rights: A Theory of Legal Responsibility,” *Yale Law Journal* 111 (2001): 443.

We offer a two-layered analysis of “corporate human rights responsibility.” The first layer of analysis explores the emergence of measurable and comparable representations and approximations of “corporate human rights responsibility.” At the first layer of analysis, we describe the rising importance of human rights indicators and their symbolic representation in maps, graphs, tables, diagrams, insignias, and social-branding labels. The second layer of analysis explores “corporate human rights responsibility” as a performance simultaneously enacted through two interacting social modalities: regions and networks. Our thesis is that indicators play a key role in sustaining and performing “corporate human rights responsibility.” We further suggest that these indicators tend to generate secondary and even third-tiered indicators (“indicators of indicators”). Ultimately, we show an epistemic (rather than merely methodological) gap between the risk to people whose human rights may be violated due to corporate activity and the set of indicators that enact and sustain “corporate human rights responsibility.”

Following the postulate that “[i]ndicators represent a distinctive method of producing knowledge about societies,”² we believe that the account that follows also has implications at the level of social theory—our second layer of analysis. We mainly draw on insights derived from Actor-Network-Theory, specifically concerning the nature of social scientific instruments such as surveys. The basic insight here is that quantitative surveys do not simply describe or capture reality (e.g., attitudes and subjects able to assume such attitudes) but also “tend to enact” these realities into being. Yet we wish to spend as little time as possible explaining “the theory” and as much time (and space) as possible describing some concrete forms and processes of enactment. In general, we treat “corporate human rights responsibility” as a definite (though certainly not finite) social formation: “it” is made of subjects and objects; “it” is made of the relations between these subjects and objects; and from our point of view “it” does not exist outside its own performances.³ It is this performativity that we are after, hoping to discover the logic of social assembly that allows “it” to perform itself as such.

We follow the postulate that “[t]he social doesn’t exist as a single spatial type. Rather, it performs several kinds of space in which different operations take place.”⁴ Applying the logic and methods of material-semiotics (also known as actor-network-theory), our premise is that by identifying different spatial modalities we may better understand how social formations are assembled and, at the level of sociological foundations, how principles of difference and similarity are enabled and performed.

Accordingly, we embed our description of “corporate human rights responsibility” within a theoretical framework of thinking about two major types of social spatialities: regions and networks. We hope to show that both regions and networks

² Kevin Davis, Benedict Kingsbury, and Sally Engle Merry, “Indicators as a Technology of Global Governance,” *Law and Society Review* 46:1 (2012): 4.

³ Our argument here is analogous to the argument that the survey creates realities “but only in the context of its own interviews.” John Law, “Seeing Like a Survey,” *Cultural Sociology* 3:2 (2009): 239, 245.

⁴ Annemarie Mol and John Law, “Regions, Networks and Fluids: Anaemia and Social Topology,” *Social Studies of Science* 24 (1994): 641, 643.

are created, sustained, and related to each other in allowing the existence of “corporate human rights responsibility.” In particular, we emphasize the extent to which the indicators we describe at the first layer of analysis are actors (“actants”),⁵ whose work simultaneously assembles the net of “corporate human rights responsibility” and the compilations that produce regions. We end up suggesting that indicators are double agents: sometimes their work produces the coherence we identify as “corporate human rights responsibility” and sometimes their work serves the purpose of distinguishing between various types and levels of such responsibility. Moreover, indicators sometimes work for the network and at other times for regions.⁶

The drive to measure

Two decades of political struggles over the meaning and scope of “corporate human rights responsibility” have culminated in the appearance of a business-case approach.⁷ Nurtured, articulated, and tested by business-management academics, this approach stipulates that the pursuit and adoption of human rights policies is not simply the morally right thing to do but also a sound business strategy.⁸ The validity of the business-case approach is often displayed in the form of graphs and diagrams showing that responsible corporations perform better than others. For example, the diagram below (Figure 5.1) compares the share-value performance of Standard & Poor’s top 500 companies and firms listed on the Financial Times Stock Exchange top 500, to an index of “most ethical corporations” (WME).⁹

The business-case model has also aligned itself with the ascendant risk-based approach to the management of firms (Enterprise Risk Management).¹⁰ Consequently various national and international bodies, strategic consultancy firms, standard-setting organizations, and business-management experts have begun to frame the human rights responsibilities of corporations in terms of risk management.

The essence of the business-case approach to human rights, writes the UN High Commissioner for Human Rights, is that business firms risk their own interests

⁵ Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (New York: Oxford University Press, 2005).

⁶ For the view that “flows are created by a subtle interplay of similarity and difference among nodes,” see Felix Stalder, “The Status of Objects in the Space of Flows,” presented at the Doors of Perception conference (Amsterdam, November 14–16, 2002), 3.

⁷ The business-case approach to human rights is a sub-stratum of a broader “business case for corporate social responsibility.” Ronen Shamir, “Capitalism, Governance, and Authority: The Case of Corporate Social Responsibility,” *Annual Review of Law and Social Science* 6 (2010): 531.

⁸ Tarek F. Maassarani, Margo T. Drakos, and Joanna Pajkowska, “Extracting Corporate Responsibility: Towards a Human Rights Impact Assessment,” *Cornell International Law Journal* 40 (2007): 135.

⁹ The WME ethical-index is compiled by the Ethisphere Institute, a commercial business-strategy player in the field of corporate social responsibility (<<http://ethisphere.com>>).

¹⁰ Michael Power, “Enterprise Risk Management and the Organization of Uncertainty in Financial Institutions,” in Karin Knorr-Cetina and Alex Preda (eds), *The Sociology of Financial Markets* (Oxford: Oxford University Press, 2005), 250.

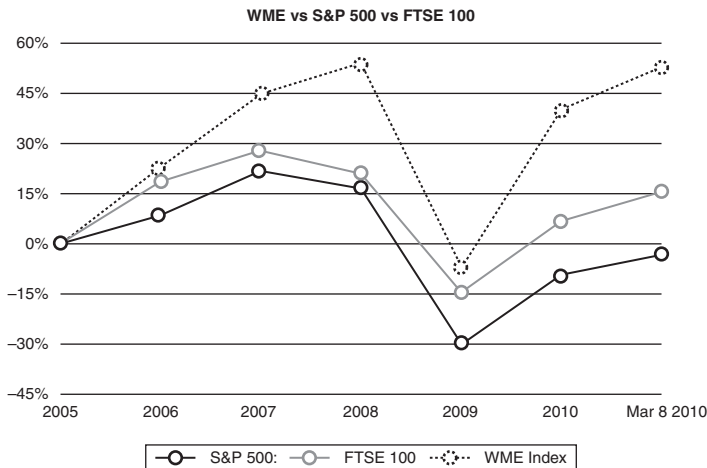


Fig. 5.1 WME v S & P

See <<http://ethisphere.com/past-wme-honorees/wme2011>>

when they fail to discover and mitigate the risks involved with human rights violations.¹¹ Another UN publication identifies three types of risk associated with a marred corporate record of human rights: reputational risks, exposure to lawsuits, and divestment. The publication also lists the business benefits that accompany respect for human rights: improved stakeholder relations, improved labor relations, a more secure social license to operate, and increased security of investments.¹² And yet another UN adviser suggests that the business-case-for-human-rights approach provides the best argument against the risk of binding and enforceable regulation. All in all the idea is that “[a]ny increased costs that may be incurred as a result of responsible human rights commitment must be seen as an ‘insurance premium’ against such risks becoming reality.”¹³ In short, we have reached a stage at which corporate concerns for human rights have become an economic risk factor which corporations must consider within their overall framework of action.¹⁴

The language of risk has become dominant in the field. Yet it is not the risk to people that the business case addresses, at least not directly. Rather it is with the risk to corporate profits—presumably posed by direct or indirect corporate involvement

¹¹ Mary Robinson, “The Business Case for Human Rights,” *Visions of Ethical Business*, in <<http://www.unhcr.ch/hurricane/hurricane.nsf/0/E47D352DEDC39697802566DE0043B28E?open=document>> (1998).

¹² Global Compact Network, “Business Case for the Human Rights Principles,” in <<http://www.globalcompact.org.ua/EN/businesscase/humanrights>> (2010).

¹³ Klaus M. Leisinger, “On Corporate Responsibility for Human Rights,” in <http://www.globalcompact.org/docs/news_events/9.6/corpresforhr_kl.pdf> (2006), 17.

¹⁴ Also see the reasoning of the Global Reporting Initiative: “Integrating human rights criteria in screening or including human rights in performance requirements can be part of a strategy to reduce the risks of investment,” in <<http://www.globalreporting.org/reporting/guidelines-online/G31Online/StandardDisclosures/HumanRights/Pages/HRIIndicatorProtocol.aspx>>.

in violations of human rights—that the business case is concerned. The underlying assumption seems to be that when a corporation follows and applies the logic of the business-case approach to human rights, it contributes to the actual protection of human rights. It seems that by protecting itself from risks to its reputation or investment a corporation also enhances the protection of potentially vulnerable populations.¹⁵ This is what we mean when we talk about a gap, or a distance, between the risk to people whose human rights may be violated due to corporate activity and the set of indicators that enact and sustain “corporate human rights responsibility.”

Davis et al. suggest that global governance indicators tacitly embody theories about the standards against which “to measure societies.”¹⁶ Along this line, we are interested in the process whereby the application of the business case by any given corporation becomes a sort of indication of its “corporate human rights responsibility.” After all, the business case for human rights cannot simply function as an abstract or principled scientific model for managing corporate risks. In order to exist and enjoy legitimacy it must meet the minimum requirement of being able to convince corporate skeptics that it indeed pays to invest in the prevention of human rights violations. In order to do this, the business case for human rights must be empirically validated. In other words, there must be data, and numbers, and measurements, and people who know how to compile, arrange, and present the data. And there must be methods and procedures that enable comparisons, and further procedures to make sure that what is compared is indeed comparable.

We focus on two initiatives that are widely recognized as the epicenter of “corporate human rights responsibility”: the UN-sponsored “Global Compact” (GC) and the “Global Reporting Initiative” (GRI).¹⁷ Both initiatives make consistent references to the value of adhering to human rights principles as an instrument for managing and reducing corporate risks. The GC asks companies to voluntarily comply, within their sphere of influence, with a set of ten core principles in the areas of human rights, labor, the environment, and anti-corruption. The first and second principles of the GC directly address corporate responsibilities to human rights. The first principle states that “[b]usinesses should support and respect the protection of internationally proclaimed human rights.” The second principle holds that “[b]usinesses should make sure they are not complicit in human rights abuses.”

The purpose of the GRI is to create a level playing field for corporations that report upon topics associated with “corporate social responsibility” (including human rights). While the GC is the most widely used vehicle for corporations who wish to assert their commitment to human rights, the GRI is fast becoming a

¹⁵ On the similar methods which are employed in the consultancy area of “impact investment” also see Sarah Dadush, “Impact Investment Indicators: A Critical Assessment,” this volume, pp. 392–434.

¹⁶ Davis, Kingsbury, and Merry, “Indicators as a Technology of Global Governance” (n.2), 74, 77.

¹⁷ Georg Kell, “The Global Compact: Origins, Operations, Progress, Challenges,” *Journal of Corporate Citizenship* 11 (2003): 35; Halina S. Brown, Martin de Jong, and Teodorina Lessidrenska, “The Rise of the Global Reporting Initiative (GRI) as a Case of Institutional Entrepreneurship,” Corporate Social Responsibility Initiative, 2007 Working Paper 36, Cambridge MA, John F. Kennedy School of Government.

global standard of reporting on such commitments. It is on the basis of their global prominence and mutually complementary functions that the GC and the GRI are increasingly cooperative. The GRI integrates compliance with the GC's ten principles into its Sustainability Reporting Guidelines. The GC, for its part, adopts the GRI's reporting guidelines as the recommended framework for its participant corporations.¹⁸ In the next sections, we describe how the GC and the GRI frame, direct, and evaluate corporate performance in the field of human rights.

Corporate risk and dangerous regions

Pledging allegiance to a publicly recognized initiative of corporate responsibility is a necessary step for any corporation that wishes to follow the dictates of the business-case approach and display a commitment to human rights. Becoming a participant in the UN Global Compact is therefore a definite move in the right direction. Any corporation on earth may join the GC and make its affiliation public by wearing the GC insignia (see Figure 5.2):

Yet, in order to be formally recognized as participants, corporations must meet basic requirements. The most important requirement is to annually submit a COP (Communication on Progress) that specifies steps the corporation has taken in support of the GC principles. In turn, the submission of a COP, and the extent to which it complies with some formal criteria, becomes an indicator of corporate support for GC principles and, by extension, of corporate commitment to human rights.

Various GC manuals advise corporations on how best to assess their standing with respect to the GC principles and how to report on it in their COPs. A primary guideline of the GC refers to the relationship between human rights risks and the country where a corporation conducts its business. It states that the danger of infringing on human rights may be greater where there is "weak governance"



Fig. 5.2 Global Compact openly accessible and explicitly free for use 'sample insignia'.

See <http://www.unglobalcompact.org/AboutTheGC/Global_Compact_Logo/GC_Logo_Policy.html>

¹⁸ <<http://www.unglobalcompact.org/news/50-06-24-2010>>.

(i.e., failed, corrupt, despotic, or otherwise abusive regimes). In order for any given corporation to assess the degree of risk associated with particular locations, the GC refers companies to organizations that collect data and develop indicators of corporate risk on a country-specific basis. In other words, corporations that rely on indicators of country-specific risks presumably make a positive step towards meeting the criteria of indicators that measure support for GC principles and therefore presumably minimize their risks (such risks being an approximation of risks-to-people).

The GC displays this sensitivity to location in the self-assessment tool that it offers to corporations so they can evaluate their standing in respect to risks. We followed the path that a corporation follows when it seeks to assess its level of risk in respect to violations of the human rights principles of the GC. The path begins at the website, <<http://www.globalcompactselfassessment.org/getstarted/followupaction>>, where, once the “human rights” area is chosen, a follow-up menu offers seven categories from which to choose. One category is “country risk” and its indicative question is: “Does the company seek to avoid involvement in human rights abuses owing to government or societal practices in its country or countries of operation?” The answer to the question functions as one indicator of corporate adherence to the human rights principles of the GC.

Respondents use four sub-indicators in order to answer the indicative question. Two of the sub-indicators assess various forms of communication between the respondent and its social environment, implicitly assuming that better communication reduces country-specific risks to the corporation. The first two indicators check whether respondents rely on country-specific corporate risk indicators. To that end, the GC singles out two organizations that specialize in assembling relevant indicators: the Danish Institute for Human Rights (DIHR) and Maplecroft. The DIHR is a government-sponsored and legally sanctioned research institution and Maplecroft is a commercial business-consultancy firm. The two specialize in corporations and offer paid services for country risk assessment in the area of human rights.

Maplecroft produces a variety of risk indices for companies, such as the Human Rights Risk Index (see Figure 5.3), generating global rankings, country reports, and a “world map” of the worst offending countries.¹⁹ The index measures 24 categories of human rights violations in areas such as security, labor, and political rights in 196 countries. The rating is based on data from bodies such as Amnesty International, Human Rights Watch, and the US State Department. The results include graphics and maps showing human rights risks to corporations on a country-by-country basis, calculated with respect to the specific risks that concern corporations.

The DIHR is more legally oriented, comparing universal standards of human rights to national laws and their level of implementation in different countries as a measure of risk to corporations. Each human right referred to in the Universal

¹⁹ <<http://www.maplecroft.com>>.

Human Rights Risk Index 2010



Legend	
Extreme risk	■ ■
High risk	■ ■
Medium risk	■ ■
Low risk	■ ■
No data	■

Rank	Country	Category
1	Afghanistan	Extreme
2	Somalia	Extreme
3	DR Congo	Extreme
4	Pakistan	Extreme
5	Iraq	Extreme

Rank	Country	Category
6	Chad	Extreme
7	Myanmar	Extreme
8	Sudan	Extreme
9	Colombia	Extreme
10	Zimbabwe	Extreme

Fig. 5.3 Human Rights Risk Index 2010

See <<http://maplecroft.com/about/news/hrra.html>>

Declaration of Human Rights and deemed relevant to corporate activities gets a three-tiered rating (“high,” “medium,” “low”) for each individual country, thereby presumably allowing corporations to assess the risk level in their operating environment. The outcome for each individual country in intersection with relevant business issues appears in Figure 5.4.²⁰

Each arrow points at a tendency concerning the relation between country and corporation, leading to a detailed table that looks, for instance in the case of South Africa, like Figure 5.5.

The accumulated results of all countries can then be represented on world maps showing degrees of (colored) risk in each country. There are multiple options of representation here. For example, it is possible to create a world map showing the relation of any specific human right in a given country to the universal standard,

²⁰ <http://www.humanrightsbusiness.org/?f=country_risk>.

RISK PROFILE

The following issues have been assigned a risk rating based on conditions in law and practice, as well as likely proximity to private sector operations. Risk ratings have been assigned on the following basis:

- ⬆ Rights issues where systematic violations place and proximity to company operations is high
- ➡ Violations take place but are not proximate to company operations
- ⬇ Rights issues where violations do not appear to take place systematically.

Fig. 5.4 This representation was taken from the open website of the Danish Institute. See <http://www.humanrightsbusiness.org/?f=country_risk> (last accessed April 20, 2010).

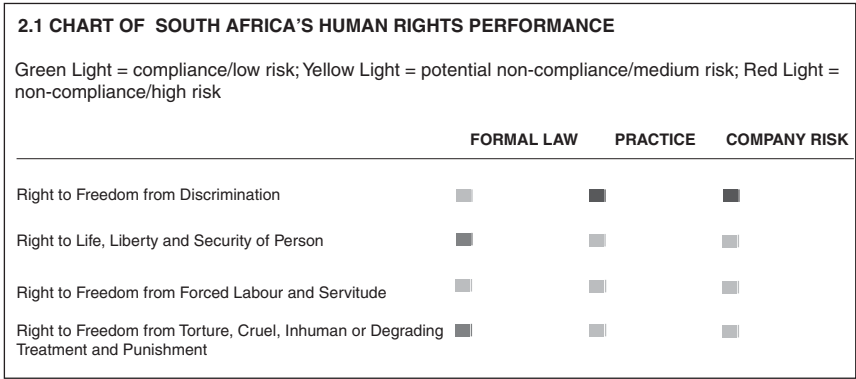


Fig. 5.5 South Africa See <http://www.humanrightsbusiness.org/?f=country_risk> (last accessed April 20, 2010).

thereby ending up with a three-color map showing the situation in each country. A corporation that subscribes to the DIHR may ask it to produce a corporate-tailored map, showing how the company is situated in respect to human rights risks in all its countries and regions of operations.

The indicators, and maps generated from the indicators, are signals of risks and degrees of risks associated with doing business in a particular locality. And as aforementioned, we are now at a point where indicators of country-specific corporate risks operate, alongside other indicators of the self-assessment tool, as means of measuring corporate variance in support of the GC principles. Indicators purport to describe a reality, but at the same time they perform and enact this reality: they enact corporations with a conscience (a form of subjectivity), which join forces to establish a recognizable social formation (i.e., corporate human rights responsibility). Yet it may well be, to paraphrase John Law, that just like surveys, compiled indicators only create a reality in the context of their own measurements.²¹ In other words, the question that remains open is whether we may identify a limited relationship between what indicators indicate and what corporations actually do.

There are four optional answers to the main question and its four indicators in the self-assessment tool: “yes,” “no,” “don’t know,” and “not applicable.” In

²¹ John Law, “Seeing Like a Survey,” *Cultural Sociology* 3:2 (2009): 239.

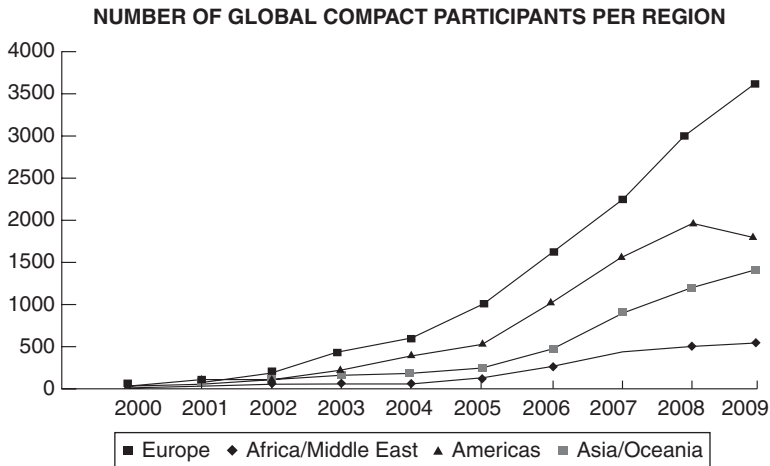


Fig. 5.6 Global Compact participants per region

See <http://www.unglobalcompact.org/docs/networks_around_world_doc/Annual_Report_2010/GCLN_2010.pdf>

order to answer the question, the self-assessor calculates a plausible average (left to the corporation's discretion) of the answers to the four indicators and then uses this average as the answer. The results are for the respondent's eyes only. The indicators only serve as an on-site learning tool about how to address the main question and, more generally, about the preparation of a Communication on Progress.²² A corporation that uses the self-assessment tool obtains an indication of its risk of falling short in its adherence to the GC principles and by extension, an indication of the risks to its own stability. The results may or may not be reported in the COP. Whether they are or not, the self-assessment mechanism illustrates the creation of indicators that move further away from the risk-to-people. The COPs which are received from participating corporations are published on the GC website (Figure 5.6).

We can see that Europe hosts a large and consistently growing number of companies that participate in the GC (submitting COPs). Africa, on the other hand, lags behind both in terms of the sheer number of corporations that participate in the GC and in terms of the modest rate of increase in corporation participation over time. Boundaries are drawn, separating regions and countries on the basis of disparate factors, such as the level of corporate risk to human rights or the number of corporations participating in the GC initiative. In this way, the diagram implies different levels of corporate protection of human rights in different countries or regions.

We have thus far focused on the generation and function of indicators and have identified two types of regional measurement: country-specific indicators of

²² John G. Ruggie, "The Theory and Practice of Learning Networks: Corporate Social Responsibility and the Global Compact," *The Journal of Corporate Citizenship* 5 (2002): 27.

corporate risk and regional-specific levels of participation in the Global Compact. We already discussed the growing distance between such indicators and the actual risk to people. We now proceed to our second layer of analysis, namely to examine the spatial types or medium through which “corporate human rights responsibility” is enabled and performed.

We are dealing here with the most common way to imagine social configurations, namely with regional spatiality (“regionalism”). Regionalism rests on the constitution of boundaries that mark and separate presumably different clusters of subjects and objects. The main sociological point about regionalism is that “when a region is defined the differences within it are suppressed.”²³ The principle of difference is created through a boundary that sets regions apart, while similarity (which signals average, common, simplified characteristics) is assumed to reign “within” each cluster. Regionalism is thus an exclusive socio-spatial modality which sustains a version of the social that performs itself through more or less neat divisions between “here” and “there.” To put it somewhat differently, one way of constructing the social (in our case, “corporate human rights responsibility”) is through symbolic and technical practices that stress similarities within boundaries and differences across them.

Yet all this becomes more complicated in the present case because multinational corporations work across boundaries and yet are expected to maintain a consistent level of responsibility to human rights wherever they operate. In this sense, regions and locations that are geographically far apart may be treated as similar or relatively identical from the point of view of measuring human rights risks. For the sake of illustration, let us assume that the country risk indicators show that the level of corporate risk is higher (and roughly similar) in both Indonesia and Colombia and lower (and roughly similar) in both Bulgaria and Morocco. In this case, regionalism in the field of human rights may not only follow the established geographical distinctions between Africa and Europe, but also generate new “regions” that spread across the globe, for example, connecting Indonesia and Colombia in a band of dark red, and linking Bulgaria and Morocco in a different “regional” band painted light yellow. In other words, we suggest that geographical distance is not necessarily what underlies regionalism. Once representational maps of similarly applied indicators are created, it is possible to generate new regions that will appear on the map not in terms of geographical location but in terms of shared human rights ratings.

The simple point is that indicators—and technologies for comparing and reporting upon them—are needed in order to create such regions and boundaries (whether old or new). This requires a network able to measure “human rights responsibility” across the globe: a network of people and methods and instruments of measurement that produce, affirm, and sustain “regions.” So regions, while real, are not a natural phenomenon residing “in the order of things.”²⁴ Rather, they are

²³ Mol and Law, “Regions, Networks, and Fluids” (n. 4), 646.

²⁴ Mol and Law, “Regions, Networks, and Fluids” (n. 4), 648.

constructs—products of another type of social formation, namely network spatiality. The next section discusses this network and what it does.

Network spatiality: A road map to corporate sustainability

As mentioned above, all GC-participating corporations must submit a COP report annually or risk being removed from the list in the event of continued non-compliance. All such reports must include an endorsing statement by the reporting corporation's CEO, a description of the firm's implementation of the GC principles, and details of the policies that the corporation has introduced in order to assess its performance and measure outcomes regarding relevant issues.²⁵ A COP need not be a specific separate document and is usually incorporated into other company texts—typically an annual Corporate Social Responsibility Report. Such documents contain statements of support for the principles of the Global Compact in radically varying degrees of elaboration. At times, a COP lists the principles of the GC and expresses its continued support for and implementation of the principles only in general terms (see Figure 5.7).

The evaluation of COPs is based on the reporting frameworks that corporations adopt in disclosing their policies and actions with respect to the GC principles. Although no one template is required, the GC advises its participants to prepare their COPs in line with the reporting guidelines and criteria of the GRI: "While the GRI G3 Guidelines do not represent the only way of preparing a COP they do offer globally recognized reporting guidance and are recommended by the Global Compact." A high-quality COP thus provides an index cross-referencing each principle of the GC to one or more corresponding GRI indicators. The table of contents of a typical report resembles the example in Figure 5.8.

Prior to their ultimate removal from the list of GC participants, corporations that fail to submit a COP are listed as "non-communicating participants." On the other end of the spectrum, COPs that meet certain reporting criteria entitle

● Support for the United Nations Global Compact

[The company] supports the 10 principles of the Global Compact with respect to human rights, labour rights, the protection of the environment and anti-corruption to actively promote our CSR activities (announced in July 2008).

We are committed to making the global compact and its principles part of the strategy, culture and day-to-day operations of our Company and to clearly stating this commitment to our employees, partners, clients and the public. Also, we espouse public accountability and transparency and will report our progress publicly.



Fig. 5.7 COP (declaration of support)

See <http://www.unglobalcompact.org/system/attachments/4678/original/COP_NEXCO-Central_.pdf?1268727027>

²⁵ <http://www.unglobalcompact.org/docs/communication_on_progress/COP_Policy.pdf>.

GRI/UNGC TABLE		Global Reporting Initiative (GRI) Indicators	United Nations Global Compact Principles
	Page		
Embracing our responsibility Overview	6		
CEO message	7	1.1	
Who we are	8	1.2, 2.1, 2.2, 2.9	
Our focus	9	1.2, AF33	1
Our history	10	2.1, 3.2	
Stakeholders	11–13	4.16, 4.17, AF12, AF14, EC2, EN26, HR6, HR7	1, 2, 3, 4, 5
Governance	14	4.3, 4.4	10
Organized for results	15	AF2, 4.1, 4.2, 4.9	1, 10
Being resourceful	16	EN26, 4.16	1, 2
Code of business conduct	17	4.6, 4.8	10
Advancing public policy	18	SO5, SO6	1
Code of vendor conduct	19	4.8, 4.12, AF1, AF2, AF3, AF6	1, 2
Product safety	20	PR2, PR4	
Goals and progress			
Supply chain Overview	22	1.2	
Over evolution	23	AF1, AF2, AF3, AF5, 4.16	1, 2
Factory working conditions	24	1.2, 4.16	2
Helping factories move forward	25	AF2, AF3, AF4, AF5	1.2

Fig. 5.8 GRI/UNGC

See <<http://www.gapinc.com/content/dam/csr/documents/SR%20Full%20Report.pdf>>

the corporation to the status of “Notable COP.” In order to be considered “notable,” a COP must set “an exceptional example” in two or more of four categories. Each category includes a number of criteria, and companies who wish their reports to be designated as notable must address a minimum number of criteria in each category.²⁶

All in all, the list of participants generated by the GC provides general information concerning the level and quality of reporting on human rights issues, roughly distinguishing between non-reporting, standard reporting, and notable reporting. In other words, reporting performance is on its way to becoming a principal indicator of “corporate human rights responsibility,” further distancing the indicators from the actual risk-to-people.

From the point of view of the social theory that we follow, the level or form of reporting is an indicator that operates through the logic of a network, which measures and compares “corporate human rights responsibility” regardless of the

²⁶ The categories are (1) a strong statement of continued support for the GC; (2) clear and detailed description of practical actions taken in implementing the GC principles; (3) measurement of outcomes that allows for checking progress; and (4) reporting process ensures reliability, clarity and timeliness of information and includes stakeholder dialogue. Using the GRI set of standard indicators is considered one of the third category’s criteria for notable COPs. There were 351 Notable COPs on the list as of August 10th, 2010. As of October 2010, the Notable Communications on Progress programme was replaced by the GC Advanced Level, <http://www.unglobalcompact.org/COP/analyzing-progress/notable_cops.html>.

geographical locations where corporations operate.²⁷ A main feature of networks is that they transcend regional boundaries and produce proximity (similarity) between locations that are—from the point of view of regional spatiality—geographically far apart (thereby relying on “flows”).²⁸ Networks depend on specific relations among their constituent elements (e.g., machines, procedures, knowledge, and skills) and it is the relative consistency of the processes of translation between them that keeps the network functioning. Networks, in the words of Latour are made of immutable mobiles with invariable links among them: similarity is a matter of relatively stable material-semiotic patterns.²⁹ We expect the assessments of corporate human rights accountability, being part of a network, to behave the same (i.e., to yield results based on similar techniques of measurement) whether they are prepared in Japan or in Saudi Arabia. Indeed, it seems that indicators are a type of actor particularly suited to sustaining networks. As Espeland and Sauder observe, the power of indicators is that numbers circulate easily and in a relatively stable manner.³⁰ Yet there are risks involved.³¹

We know that there is significant variance in the type of data that corporations collect and measure and considerable variation in the way they communicate this data to the network. When assessments vary in their measurement techniques, address different and selective issues, or simply are poorly conducted in one place due to lack of data or skills, then both mutation and variance emerge, leading to network failure. The greater the variance in assessment data, the more the network is prone to fail in its diagnosis and measurement of deviations from the human rights standards that it seeks to uphold. The particular social formation that the network performs—in this case corporate human rights accountability—begins to dissolve.

It is precisely the danger of network failure that the GRI is designed to resolve. The GRI aims to create an immutable mobile: a technology of reporting based on 79 “performance indicators” (49 “core” and 30 “additional”) in six categories of corporate social responsibility (human rights being one of them). This technology is supposed to generate consistency and stability of measurement wherever it

²⁷ On networks as forms of governance, see Davis, Kingsbury, and Merry, “Indicators as a Technology of Global Governance” (n. 2), 82, 88.

²⁸ Manuel Castells, *The Rise of the Network Society, the Information Age: Economy, Society, and Culture* Vol. 1 (Cambridge: Oxford University Press, 1996).

²⁹ See Bruno Latour, *Science in Action: How to follow scientists and engineers through society* (Cambridge: Cambridge University Press, 1987). See also Michael J. Gorman, “The Elusive Origins of the Immutable Mobile,” <<http://www.stanford.edu/group/STS/immutablemobile.htm>>, 2001.

³⁰ Wendy Nelson Espeland and Michael Sauder, “The Dynamism of Indicators,” this volume, pp. 86–109.

³¹ For the fundamental transformation of information contained in standard questionnaires see Ronen Shamir and Dafna Hacker, “Colonialism’s Civilizing Mission: The Case of the Indian Hemp Drug Commission,” *Law and Social Inquiry* 26:2 (2001): 435. The authors analyze the late 19th-century report of the Indian Hemp Drug Royal Commission of Inquiry. The Commission found that the British scientific belief in the connection between the consumption of hemp products and mental insanity originated in an over-zealous adherence to standard forms. Standard questionnaires required asylum personnel to always fill the “cause of insanity” entry of the forms. These forms then travelled from each asylum to New Delhi, where a further compilation had been aggregated and sent to British scientists in London. In the process, the original intent in filling the forms had been transformed into authoritative statistical data concerning the impact of hemp drugs on one’s mental health.

operates and regardless of the identity of the corporation that follows its template. The GRI regularly publishes a list of corporations that adhere to its reporting guidelines. In order to appear on the list, a company must specify a “content index” of the GRI indicators that it addresses, cross-reference these indicators to other relevant company documents, and explain instances when it fails to address (in part or in full) some of these indicators.

The GRI reporting guidelines refer to nine indicators in the category of human rights, six of which are classified as “core” indicators and three as “additional” indicators. The core indicators ask corporations to report on issues such as: the number of its business agreements that include human rights clauses; the number of its contractors that have undergone human rights-related screening; instances of discrimination cases and collective bargaining infringements; and policies and practices concerning child and forced labor. Three additional indicators address the percentage of corporate personnel in general and security personnel in particular who are trained in the firm’s human rights policies and procedures, and the number of incidents involving human rights violations of indigenous populations.

Still, the GRI, like all other instruments in this network, is not a mandatory and enforceable device. Therefore, even companies that follow the reporting guidelines of the GRI are not required to address each and every one of the GRI’s indicators or to follow all its articles of disclosure. In order to address this risk of network failure, the GRI requires corporations to measure and report upon their level and scope of reporting. To this end, the GRI developed another set of indicators that measure a corporation’s “application level.”

A corporation that seeks to be listed must declare its “level of application.” This is a quantitative criterion that measures how many of the GRI indicators the company addresses. There are three “levels of application” according to which corporations are classified: A, B, and C. To earn a “C,” a report must address at least ten “performance indicators” including at least one belonging to the “economic” category, one to the “environmental” category, and one indicator relating to one of four other categories (labor, human rights, society,³² and product responsibility). In practice, a “C” level report may therefore avoid addressing any human rights-related indicator. A level “B” report should address at least twenty “performance indicators” including at least one of each category (including human rights). A level “A” report must address all 49 core performance indicators and this means that such corporations report on all of the six core human rights principles.

There is an additional factor according to which corporations may be distinguished with respect to their application of the GRI. A report that underwent external auditing (e.g., was certified by an accounting firm like PricewaterhouseCoopers) is also graded a “+” alongside its “A,” “B,” or “C” level. A third method of differentiation among reporting corporations concerns their “declaration status.” Each of the “A,” “B,” or “C” levels of reports may be self-declared by the corporation, affirmed by an independent third party, or audited by the GRI (a paid

³² The category of “society” lumps together issues such as “community,” “corruption,” and “anti-competitive” behavior.



Fig. 5.9 GRI self-declared Application Level icons

See <<https://www.globalreporting.org/reporting/reporting-framework-overview/application-level-information/Pages/default.aspx>>

service). The highest quality of reporting thereby entitles a corporation to use an A+ insignia that also bears the words “GRI-checked.” The lowest level of reporting along that scale is therefore a “self-declared” “C.” Each such insignia has its own color (see Figure 5.9).³³

This set of indicators therefore seeks to create a comparable standard of reporting across the network. Standardized variance in the “level of application” in turn becomes an indicator of the indicators of commitment to human rights responsibility (our first layer of analysis). The principle of difference is not based on boundaries between regions but on different levels of reporting that, in turn, enable a hierarchy of reporting corporations as a measure of their commitment to human rights.

Yet the GRI also acknowledges network difficulties and failures. On the one hand, it celebrates “a truly global phenomenon with GRI-based reports coming from over 60 countries around the world and a global network focused on the vision of creating a set of Guidelines that will work for any organization, anywhere in the world.” On the other hand, the GRI is aware of the fact that “sustainability is very much about a time and a place and a community.” A set of problems must be dealt with: how to relate the GRI Guidelines to local policies and regulations; how to account for cultural differences; how to assess performance in the context of particular places and locations. By way of addressing such problems, the GRI is developing a system of National Annexes so as “to create a space within the GRI portfolio for reporting which is simultaneously both global and local.”³⁴ The basic idea is that alongside the universal reporting guidelines of the GRI, corporations will also prepare annexes that will address country-specific or region-specific

³³ As of January 2011 GRI no longer provides self-declared Application Level icons, but encourages users to mention the Application Level in their reports without an icon. See <<https://www.globalreporting.org/trademarks-and-copyright/Pages/Frequently-Asked-Questions.aspx>>.

³⁴ <<http://www.globalreporting.org/reporting/reporting-framework-overview/national-annexes/pages/default.aspx>>.

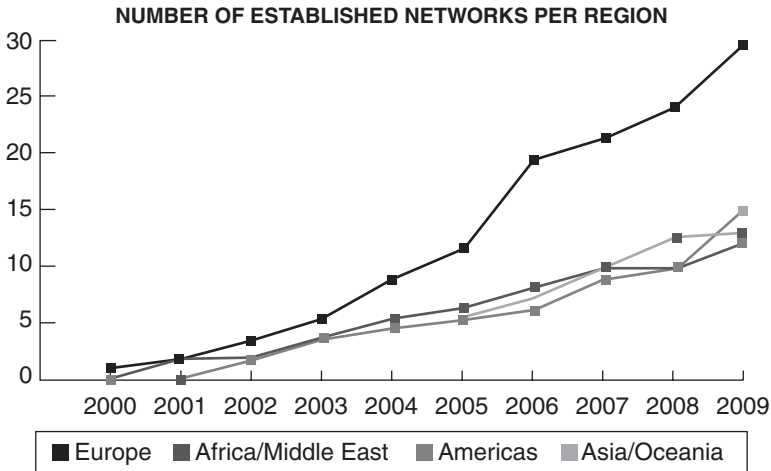


Fig. 5.10 Global Compact networks per region

See <http://www.unglobalcompact.org/docs/networks_around_world_doc/Annual_Report_2010/GCLN_2010.pdf>

sustainability issues (such as human rights) directly related to the locations of their operations.

So we are back with the regions and geographical distinctions with which we opened our discussion. The Global Compact's "local networks" project provides a strong example of the constant movement between regional and network spatialities. This project encourages GC participants—upon signing a special agreement with the GC headquarters in New York—to form networks of communication and exchange on a regional basis. The GC, for its part, measures the global spread of local networks on the basis of its standardized regional division of the world to four areas: Africa/Middle East, Americas, Asia/Oceania, and Europe/Commonwealth of Independent States. Again, as illustrated in Figure 5.10, differences are suppressed within regions and emphasized across boundaries.

The local networks function as branches of GC participants according to their geographical location. The rationale for their operation is to facilitate the creation of regional Knowledge Sharing Systems that will address "problems unique to their parts of the world."³⁵ The GC also allows regional variation in the methods of reporting: it offers country-specific COP templates alongside its recommended GRI format.³⁶ In their developed form, these local networks are also encouraged to generate—through processes of mutual learning—their own methods of advancing

³⁵ Local Networks Report, p. 6, at <http://www.unglobalcompact.org/docs/networks_around_world_doc/Annual_Report_2010/GCLN_2010.pdf>.

³⁶ Country-specific COP templates may be removed in tandem with the increased convergence between the GC and the GRI, the latter now in the process of offering its country-specific appendices.

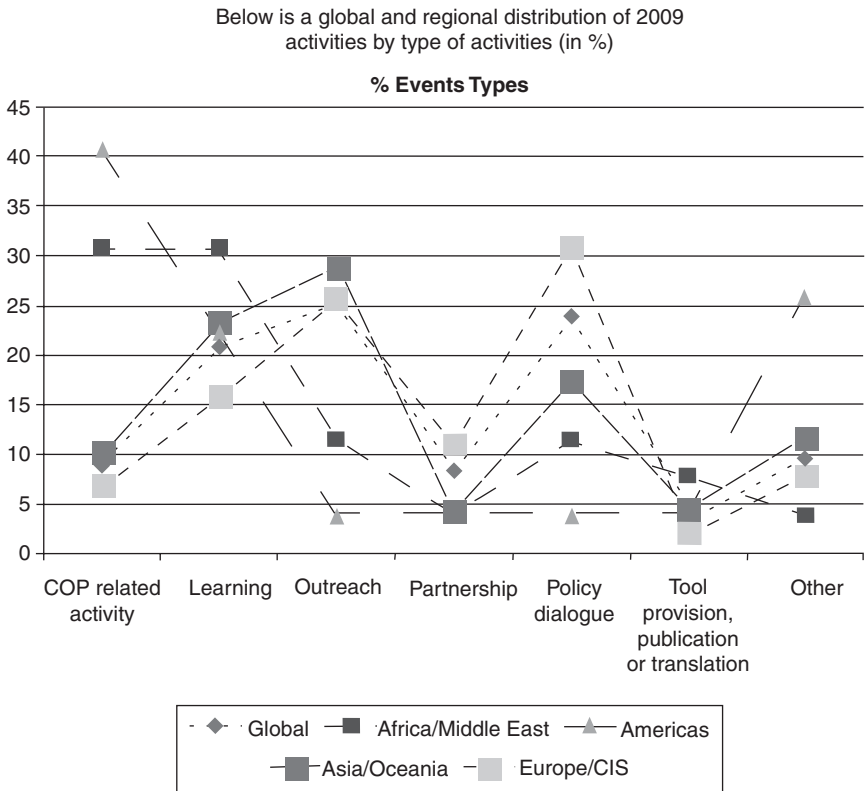


Fig. 5.11 Global Compact % events

See <http://www.unglobalcompact.org/docs/networks_around_world_doc/communication/network_reports/2009/CH_2009.pdf>

compliance with GC principles. In other words, the local networks may create their own tailor-made regional indicators regarding the scope and pace of implementation of GC principles. Moreover, we can trace the evolution of indicators that measure and compare the activities of local networks (e.g., conferences, workshops, awareness projects, etc.) against other local networks. All this is arranged and graphically displayed on a regional basis (see Figure 5.11).

The Global Compact network, through the creation of local networks, (re-)generates regionalism. The GC publishes a world map that further distinguishes between “local networks” and “emerging networks” (in the process of becoming formally recognized “local” ones) (see Figure 5.12).

The 2010 GC report on local networks includes statistical data and indicators that compare the networks’ performance on the GC standards across four geographic regions: Africa/Middle East, Americas, Asia/Oceania, and Europe/Commonwealth of Independent States.³⁷ These compiled data and comparisons

³⁷ Local Networks Report, <http://www.unglobalcompact.org/docs/networks_around_world_doc/Annual_Report_2010/GCLN_2010.pdf>.

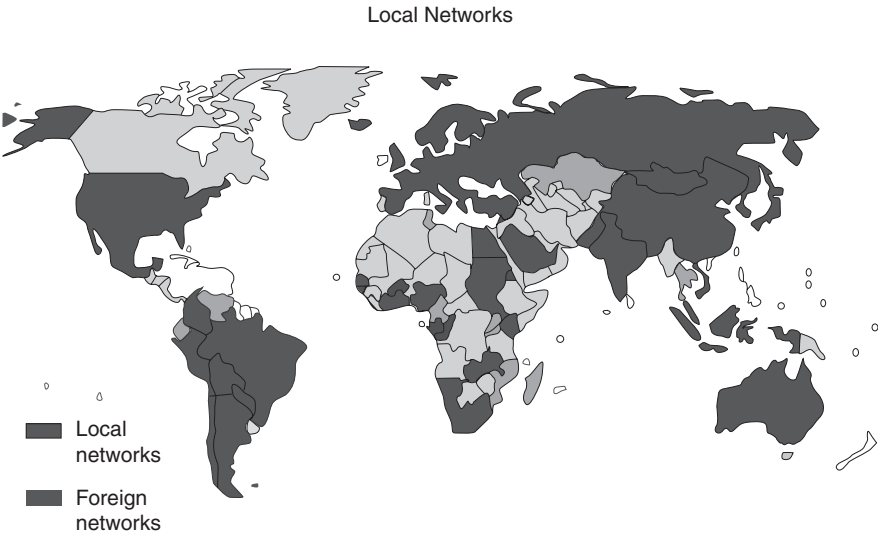


Fig. 5.12 Global Compact World Map

See <<http://www.unglobalcompact.org/NetworksAroundTheWorld/index.html>>

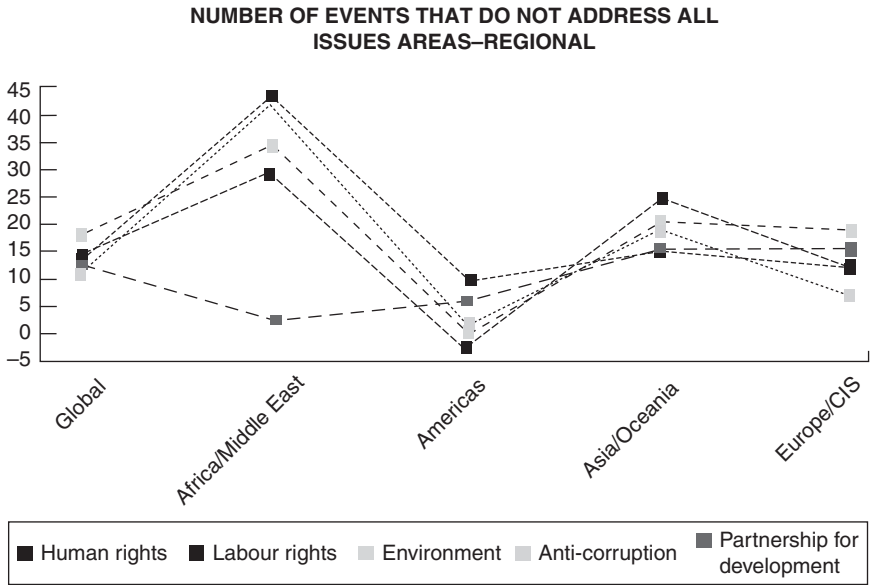


Fig. 5.13 Global Compact events

See <http://www.unglobalcompact.org/docs/networks_around_world_doc/Annual_Report_2010/GCLN_2010.pdf>

indicate that local networks in Africa raise more human rights issues in their joint meetings than their European/CIS counterparts. The data also enable comparisons across regions of the number of events dedicated to human rights compared to other issues (see Figure 5.13).

These data, which return the focus to geographic regions, support our conclusion that “corporate human rights responsibility” is a social formation that coexists in two types of spatiality. The perceived failures in each type of spatiality trigger the production of the other. This back-and-forth movement and continual development of indicators of indicators further distances the purported measurements of human rights responsibility from risk-to-people.

Summary

Throughout this chapter we have deployed two layers of analysis to tell a story about “corporate human rights responsibility.” The first layer was about the emergence of measurable and comparable representations, approximations, and indicators of “corporate human rights responsibility.” The second layer was about “corporate human rights responsibility” as a performance that is enacted simultaneously through two social modalities: regions and networks.

Concerning the first layer, our findings indicate a growing distance between the problem of corporate human rights responsibility and the methods developed for diagnosing and treating it. The problem with which we began was the risk that corporate activities may directly or indirectly violate the human rights of people within the corporation’s sphere of influence (the “risk-to-people”). A conceptual solution emerged in response to this fundamental risk—namely, the idea that corporations that do not strive to protect the human rights of potentially vulnerable people jeopardize their own sustainability (long-term profits). The move from “risk-to-people” to “risk-to-corporations” already represents one move away from directly treating the problem. Most, if not all, of the procedures that we encountered in our analysis orient themselves towards minimizing corporate risks and not “risk-to-people.” This disconnect is symptomatic of a common problem with indicators: “Indicators often express ideologies about the ideal society and the process of achieving it. But what they actually communicate, and to whom, may not be what their producers and promulgators sought to communicate.”³⁸

We focused on the Global Compact as a primary institutional vehicle through which to address corporate risk. By merely pledging support to the human rights principles of the GC, corporate participants are considered to be taking a vital preliminary step in minimizing their risk. In other words, declared support of the GC principles becomes, in and of itself, an indicator of “corporate human rights responsibility,” now already twice removed from the “risk-to-people.” We have then shown how yet another indicator is set into action: a measurement that

³⁸ Davis, Kingsbury, and Merry, “Indicators as a Technology of Global Governance” (n. 2), 78.

differentiates corporations on the basis of their compliance with their pledges to support the principles of the GC. The instrument assesses corporate reports that convey the scope of form of corporate adherence to the principles of the GC (COPs: Communication-on-Progress). The GC classified the COPs roughly into three categories of reporting. The reports that enjoy the highest esteem are those that follow the GRI reporting guidelines. The reporting framework of the GRI, in concert with the GC principles, is based on indicators that assess the human rights risks posed to people by corporate policies and practices. Yet we have shown that it is the variation between corporations in their methods of reporting, and not the substantive content of their reports, which acts as an indicator of corporate risk. We have therefore reached a stage whereby variation in the form ("level") of reporting acts as a primary indicator of "corporate human rights responsibility."

The second layer of our analysis concerned the enactment of "corporate human rights responsibility" through the social modalities of regions and networks. In the first part of the chapter we identified an emphasis on country-specific human rights risks to corporations. We found that country-specific risks (themselves rated on the basis of various indicators) function as indicators of the risks to corporations that fall behind in their support of the GC principles. As such, these indicators play their part in the distancing of "corporate human rights responsibility" from the "risk-to-people" (our first layer of analysis). Simply put, the idea (together with the measurement practices that express it) is that there are greater risks to corporations in countries where there is greater risk to human rights due to local conditions. The implication is that corporations that operate in high-risk areas must introduce stronger measures of "corporate human rights responsibility" in order to reduce their risk. At the level of social theory, therefore, the reliance on country-specific risk indicators as a measure of corporate risk embeds "corporate human rights responsibility" in a regional framework—one that hinges on distinctions between "here" and "there" and the drawing of boundaries as a constitutive principle of producing difference between regions and similarities within each of them. We also discussed the subsequent development of various indicators of indicators, for instance measurements of the number of GC participants in each world region, further contributing to the "distancing" process that we identified.

At the same time, we suggested that country-specific and regional indicators of "corporate human rights responsibility," however removed from "risk-to-people," are products of a network. This network is held together, assembled by, and enacted through reports, measurements, assessment tools, institutional actors, symbols, and the instrumentality of indicators. It is a network of "corporate human rights responsibility": a material-semiotic patterned measurement of the risk that corporate activity may directly or indirectly violate the human rights of people in their sphere of influence.

We found that efforts to create a stable pattern of comparable measurements of "corporate human rights responsibility" across the network coalesced in the creation of indicators of levels of reporting, increasing further the distance between the technology of treatment (the indicators themselves) and the effort to minimize the "risk-to-people." Yet, from the point of view of social theory, the interesting finding

here is that the inability of these “reporting-level” indicators to sustain a stable pattern of connections between the nodes on the network prompts the emphasis on the regional modality. We have seen that both the GC and the GRI have begun to encourage the production of “local networks” and country-specific indicators of risk beyond those directly related to the country-specific risks with which we began our journey. New indicators appear, this time measuring the activities of these networks and comparing them across regions. We are looking at the interplay between a regional formation and network formation. We are looking at indicators that at times work to preserve the integrity of the network and at times to enable the production of regions. Indicators are double agents that enable “corporate human rights responsibility” as a social configuration, which feeds off, and is sustained by, two interacting spatial types. They are also double-agents because they purport to address the problem of corporate-related human rights risk but in fact push “corporate human rights responsibility” further and further away from solving the risk-to-people.

6

Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order*

Nehal Bhuta

“I have simply wanted to show some sides or edges of what we could call the practico-reflexive prism or just simply the reflexive prism, in which the problem of the state appeared in the sixteenth century . . . It is a bit as if I were to say to you: my aim has not been to give you the history of the planet Earth in terms of astrophysics, but to give you the history of the reflexive prism that, at a certain moment, allowed one to think that the Earth was a planet.”¹

The emergence of state fragility

In a subterranean food court under one of Washington’s neoclassical federal buildings, an official with USAID pushed an orange internal mail envelope across the table. Opening it, the words “sensitive but unclassified” appeared at the top of the enclosed document, which was entitled “USAID 2010 Alert Lists—Fragility and Risk for Instability: World Wide Rankings.” The list ranked 161 states in an Index of Fragility, from Afghanistan in first place, to Sweden at 161st. “We have,” said the official, “been making lists for years.”

We had been discussing the proliferation since 2001 of the notion of “fragile state” in the policy language of most Western governments’ donor agencies, and adoption of the concept by the OECD’s Development Assistance Committee in 2005. It was widely understood that the attacks of September 11 had given a new lease of life to the 1990s problematic of “failed” and “failing” states (also variously

* Early and partial drafts of this chapter were presented to workshops at NYU Law School, the Buffet Center for International Studies at Northwestern University and the Yale Law School’s Human Rights Workshop. Comments from participants in those events have greatly assisted me. My thanks to Benedict Kingsbury, Sally Engle Merry, Kevin E. Davis, Ian Hurd, James Silk, and Paul Kahn for inviting me to present. Invaluable research assistance was provided by Clare Watson.

¹ Michel Foucault, *Security, Territory, Population: Lectures at the College de France, 1976–1977* (Houndsmills: Palgrave, 2007), 276.

described as “weak” or “fragile”), having dramatically illustrated the risks that could emerge from territories in which a government’s power to control its territory and population was limited.² State-funded development agencies—themselves increasingly involved throughout the 1990s in funding governance reforms and “institution-building” in the context of post-conflict reconstruction and peace-building³—quickly adopted the notion of “fragile state” as a term to describe a broad-ranging development-programming agenda that recast development assistance as a means to promote political stability and peace, thus serving a national security interest in counteracting state fragility.⁴ “Effective and legitimate” states had become an objective of development programming, which sought to address a sprawling basket of social, economic, and political arrangements within “fragile states” in order to target “the causes of fragility and vulnerability.”⁵ Development promotion and the containment of security threats were closely intertwined in this understanding.⁶

² See National Security Council, National Security Strategy of the United States, September 2002, 1; National Security Council, National Security Strategy of the United States, March 2006, 37, 44; National Security Council, National Security Strategy of the United States, May 2010, 8, 11, 13.

³ See, for example, Diana Cammack, Dinah McLeod, Alina Rocha Menocal, and Karin Christensen, “Donors and the ‘Fragile States’ Agenda: A Survey of Current Thinking and Practice,” Overseas Development Institute for the Japanese International Cooperation Agenda, March 2006. See also Stewart Patrick and Kaysie Brown, “Greater than the Sum of Its Parts? Assessing Whole of Government Approaches to Fragile States,” International Peace Academy, June 2007, available at <<http://www.cgdev.org/content/publications/detail/13893>>. A review of the websites of the UK’s Department for International Development, Germany’s Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), UNDP, Canada’s International Development Agency (CIDA), USAID, and EuropeAid shows that all of these agencies consider one or more of “governance reform,” “effective institutions,” “democratization,” and “rule of law” as programmatic priorities.

⁴ A 2004 USAID White Paper asserted: “Failed states and complex emergencies now occupy center screen among the nation’s...national security officials...Development is now *as essential to U.S. national security as are diplomacy and defence*.” USAID, “U.S. Foreign Aid: Meeting the Challenges of the Twenty First Century,” Bureau for Policy and Program Coordination US Agency for International Development, White Paper, 2004, 3, 5, 7. As one foreign policy expert put it, fragility was a way of “getting militaries and foreign ministries interested in poverty and development.” Interview with foreign policy expert, Washington DC, July 22, 2010.

⁵ David Carment, Stewart Prest, and Yiagadeesen Samy, “Determinants of State Fragility and Implications for Aid Allocation,” UNU WIDER, Research Paper No. 2008/46, April 2008, 2–3; see also USAID, “Fragile States Strategy,” Bureau for Policy and Program Coordination US Agency for International Development, January 2005; Claire Vallings and M. Moreno-Torres, “Drivers of Fragility: What Makes States Fragile?,” Department of International Development (DFID), PRDE Working Paper, April 2005.

⁶ See Stewart Patrick and Kaysie Brown, “The Pentagon and Global Development: Making Sense of the DoD’s Expanding Role,” Center of Global Development, Working Paper No. 131, 2007, noting the emphasis on addressing fragile states as part of the newly established AFRICOM’s objectives. See also United States Department of Defence, “Quadrennial Defence Review Report,” February 2006, 22, available at <<http://www.defense.gov/qdr/report/report20060203.pdf>>. For an account of factors influencing the “securitization of development,” see Bjorn Hettne, “Development and Security: Origins of the Future,” *Security Dialogue* 41 (2010): 31. The US Department of Defence’s 2008 Operations Manual on Stability Operations declares “state fragility and instability... a grave threat to national security.” Department of Defence, “Field Manual 3–07: Stability Operations,” October 2008, <<http://usacac.army.mil/cac2/repository/FM307/FM3-07.pdf>>, 1–3.

Despite—or perhaps because of—a lack of consensus⁷ on the meaning and determinants of “state fragility,” the concept has been described as “one of the most successful and influential development policy euphemisms of the past ten years.”⁸ But the murkiness of the concept, and the proliferation of competing definitions and typologies,⁹ obscure an underlying point of convergence in the widespread adoption and use of “state fragility”—that state institutions, state policies, and the relationship between state and society (variously described as a “social contract,” “state-society expectations,” “state structure,” or “constitutional order”) are fundamental “drivers” of social, economic, and political outcomes within the territory. DFID, for example, defines fragility as an incapacity or unwillingness of the state to provide essential services, reduce poverty, and settle conflicts, and identifies “weak state institutions” as “the central driver of fragility.”¹⁰ USAID declares state fragility to be the “product of ineffective and illegitimate governance,” and defines “effectiveness” as “the capability of the government to work with society to ensure the provision of order and public goods and services,” and legitimacy as “the perception by important segments of society that the government is exercising state power in ways that are . . . fair and in the interests of the nation as a whole.”¹¹ The state, it would seem, is not only back in, but more important than ever as an objective of Western developmental ambition. As appears from these definitions, the concept of state fragility is contrastive—a fragile state is one which is *not* effective, capable, legitimate, or stable. The definitions mix propositions about causes, correlates, and consequences, but it should be evident that that concept is deeply normative, ascertainable in practice only against some real or ideal type of what an effective, capable, legitimate, and stable state looks like. Otherwise disparate events, descriptions, statistics, and claims pertaining to a territory become a unified object called “fragility.”

⁷ Charles Call, “The Fallacy of the ‘Failed State’,” *Third World Quarterly* 29 (2008): 1491–507; Jorg Faust, Jorn Gravingholt, and Sebastian Zaja, “The Fragile Consensus on State Fragility: On the Difficulties of Donor Coordination,” German Development Institute (DIE), 2009, 1, noting that “there is no generally accepted concept of what constitutes fragility and what could or should be done about it.”

⁸ James Fearon, “Do Governance Indicators Predict Anything?” World Development Report, Background Paper, May 24, 2010, 1.

⁹ Department for International Development, “Why We Need to Work More Effectively in Fragile States,” DFID, January 2005, 15; USAID, “Fragile States Strategy” (n. 5), 3; OECD, “Concepts and Dilemmas of State Building in Fragile Situations: From Fragility to Resilience,” OECD Development Assistance Committee Discussion paper, 2008, 11–12; World Bank (International Development Association), “Operational Approaches and Financing in Fragile States,” June 2007, 2.

¹⁰ Vallings and Moreno-Torres, “Drivers of Fragility” (n. 5), 7. See also DFID, “Why We Need to Work More Effectively in Fragile States” (n. 9), 15: “Effective states depend on effective political leadership equipped with the skills to manage conflicting interests, agree effective policies, and see through structural change.”

¹¹ USAID, “Fragile States Strategy” (n. 5), 3. The OECD definition combines aspects of both DFID and USAID’s definition, while notable non-governmental policy definitions emphasize the lack of “effectiveness, authority and legitimacy” as determinants of state fragility: see, for example, Carment, Prest, and Samy, “Determinants of State Fragility and Implications for Aid Allocation” (n. 5); and Frances Stewart and Graham Brown, “Fragile States,” LSE CRISE, Working Paper No. 51, January 2009.

Definitional difficulties, and the heavily judgment-laden nature of the term “fragility,” ensured that diplomatic corps in many countries were reluctant to adopt and apply it. In both the Foreign and Commonwealth Office (UK) and the State Department, desk officers rejected the application of the “fragile” label to states they monitored.¹² But the widespread circulation of the term in development discourse ensured that the category remained in use, adopted by the World Bank, OECD, USAID, UK’s DFID, Australia’s AUSAID, Germany’s BMZ and GTZ, and a number of university-based research institutes.

The term’s proliferation and deployment to plan and describe strategies of development assistance, or evaluate the political, social, and economic conditions in developing states, reinforces the sense in which the concept achieves a certain “proof-in-use.”¹³ Its relationship to an underlying “object” with definite characteristics, propensities, qualities, and incidents may be difficult to pin down—indeed, as one survey of the “fragile states agenda” concluded, “fragile” is often substituted without a clear change in meaning by “failed,” “failing,” “crisis,” “weak,” “poorly performing,” “under stress,” “post-conflict,” or “at risk of instability.”¹⁴ Yet, like so much of development discourse, the term can be understood as a “trading language” used to talk about complex social realities which are highly heterogeneous and about which there is little theoretical and empirical agreement. Once the concept is pragmatically and provisionally in use—and used in a variety of diagnostic, prescriptive, and evaluative ways—its consistency with any “outer reality” is less significant for its users than its correspondence with a (porous) set of acceptable uses and deployments.¹⁵ In this sense, to recall a well-worn observation from actor-network-theory, a constructed knowledge-object is no less real than a “natural” one—its reality and the effects of its realness emerge in its iteration and use in various forms of classification and calculation.¹⁶

¹² See Patrick and Brown, “Greater than the Sum of its Parts?” (n. 3), 16, 37.

¹³ I borrow this term from: Alain Desrosières, “How Real Are Statistics? Four Possible Attitudes,” *Social Research* 68 (2001): 339–55.

¹⁴ Cammack et al., “Donors and the ‘Fragile States’ Agenda” (n. 3), 16.

¹⁵ See Richard Rottenburg, “Code-Switching, or Why a Metacode is Good to Have,” in Barbara Czarniawska and Guje Sevon (eds), *Global Ideas: How Ideas, Objects and Practices Travel in the Global Economy* (Liber AB and Copenhagen: Copenhagen Business School Press, 2005), 259–75. See also Richard Rottenburg, *Far-Fetched Facts: A Parable of Development Aid* (Cambridge, MA and London: MIT Press, 2009), 180–91.

¹⁶ See, helpfully, Donald MacKenzie, *Material Markets: How Economic Agents are Constructed* (Oxford: Oxford University Press, 2009), 22–4, 167–8; Timothy Mitchell, “The Properties of Markets,” in Donald MacKenzie, Fabian Muniesa, and Lucia Siu (eds), *Do Economists Make Markets? On the Performativity of Economics* (Princeton: Princeton University Press, 2007), 244–75; Bruno Latour, *Reassembling the Social: An Introduction to Actor Network Theory* (Oxford: Oxford University Press, 2005), 89, 118: “Artificiality and reality marching in step . . . There is no rear-world behind to be used as a judge of this one, but in this lowly world there lie in wait many more worlds that aspire to become one—or not, depending on the assembly work will be able to achieve.” See also Susan Leigh Starr and Geoffrey Bowker, *Sorting Things Out: Classification and its Consequences* (Cambridge, MA: MIT Press, 2000), 298–300. See also Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, MA: Harvard University Press, 1998), 333: “An object’s reality depends on the extent and robustness of the broadest network of the objects in which it is inscribed.”

To criticize the notion of state fragility as lacking analytical rigor or clear boundaries is in a certain sense to misunderstand how it acts: it provides a heuristic and vocabulary for organizing, describing, interpreting, and classifying, which then becomes a basis for acting in and on a complex reality. "Fragility" is thus best understood as an "assemblage" which brings into relation with one another a range of empirical claims, theories, concepts, and data that originated from diverse research agendas and institutional incubators: political science, economics, development studies, security studies, embedded in task forces, research centers, development banks, and state development agencies, to name a few. Fragility is not a "master concept" in the sense of unifying each of these dimensions or theories under a single logic; rather, it associates a range of political/social phenomena, theoretical claims, normative models, statistical inferences, and datasets and bundles them into a factual quality called "state-ness." This quality has a "thingness" to it in as much as it is considered sufficiently solid to ground consequential causal claims and inferences among those who work with the concept. For example, Carment, Prest, and Samy,¹⁷ introduce their paper on "Determinants of State Fragility and Implications for Aid Allocation," by stating that

we *know* that fragile and failed states constitute a security risk in a number of important ways . . . They are a risk to their own people because they lack capacity, resulting in a lack of basic security. They lack governance resulting in inefficient and inequitable distribution of resources; and they lack control over violence within their territory, resulting in further divisions and weakness, and the diffusion of conflict into new jurisdictions.¹⁸

Carment and his co-authors admit that states need not exhibit all of these characteristics in order to be considered "fragile," but rather contend that fragility is a "spectrum" along which states can be placed as a matter of degree.

Used in this way—and this is rather typical of the ways in which fragility is discussed¹⁹—the classification "fragile" ties the state so classified into a set of beliefs, narratives, common-sense notions, and organizational routines not only around the notion of "fragile state" but also of "stable" or "resilient" state (or whatever counter-term one might imagine). The spectrum lies in a problem-space populated by fragile and durable states (effective/ineffective states), in which the qualities of highly heterogeneous political and social orders can be mapped, grasped, known, compared, and addressed. In this way, it seems to me that the fragility category opens a path to—perhaps even promises—technically rendered knowledge of the reality of a given state and society, as a means to addressing insecurity and instability.

¹⁷ Carment leads the "Country Indicators for Foreign Policy" project at the Patterson School of International Affairs in Ottawa. Among the indexes produced by this project is an Index of State Fragility and Failure, last tabulated in 2007: <<http://www4.carleton.ca/cifp/>>.

¹⁸ Carment et al., "Determinants of State Fragility and Implications for Aid Allocation" (n. 11), 1.

¹⁹ See, e.g., Stewart Patrick and Susan Rice, "Index of State Weakness in the Developing World," Brookings Institution, February 2008; Patrick and Brown, "Greater than the Sum of Its Parts?" (n. 3); USAID, "Fragile States Strategy" (n. 5); Stewart and Brown, "Fragile States" (n. 11).

In Foucault's *dispositif* (apparatus) of security, grasping something at the level of its effective reality is a defining feature of the mechanism of security; in order to contain risks and protect well-being, a risky reality must be diagnosed and responded to by "getting the components of reality to work in relation to each other, thanks to and through a series of analyses and specific arrangements."²⁰ A mechanism of security relies on decomposable knowledge of reality, to better understand how one set of possibilities could be used to contain another, or how certain arrangements could be rearranged to counteract or nullify risks inherent in them.

"Fragility" revolves around a set of claims about how institutions, state structures, and political orders lack or obtain "legitimacy" and "effectiveness" (or willingness and capacity), and another set of claims about how state and society interact in order to generate stability or instability. Building resilience, or counter-acting fragility, as a development agenda requires—at least as a boundary presupposition—the possibility that knowledge generated by these various claims can be acted on to either strengthen the state or reorganize its relationship with society;²¹ as such, it seems to fit with Foucault's notion of a *dispositif* of security. A "Strategy Framework" developed in relation to USAID's Fragile States initiative²² expressly analogizes fragility to a medical syndrome which must be understood in its causal determinants in order to be treated:

Just as medical science develops drugs and treatments for a disease and not for a particular individual, our approach targets common sets of maladaptive institutional characteristics and their dynamics... We call these sets of maladaptive cases fragile state syndromes. Following the medical metaphor—appropriate since medicine must also address an organic system to be successful—syndromes are brief vignettes capturing the common essence of the political, economic, social and security dynamics being observed for a group of countries.²³

The Strategy Framework holds out the possibility that the right understanding of fragility can provide a "holistic framework for monitoring, analyzing and responding to the factors that underlie threats to social order . . . [and] . . . a causal analysis of how legitimacy and effectiveness can be restored." Similarly, DFID's working paper on "Drivers of Fragility" claims that the causes of fragility can be identified as "political institutions [which] are not strong enough to manage effectively the natural conflicts that occur in society," and so anticipates that "external actors can help shape institutions [in] fragile states."²⁴ Carment et al. argue that better

²⁰ Foucault, *Security, Territory, Population*, (n. 1), 47.

²¹ Rottenburg points out that all development discourses are based on the assumption that "historical changes in the social, political, legal, economic and technical capacities of human collectivities can be designed and implemented at will according to what is considered desirable." Richard Rottenburg, "Accountability for Development Aid," in Herbert Kalthoff, Richard Rottenburg, and Hans Jurgen Wagener (eds), *Ökonomie und Gesellschaft: Facts and Figures, Economic Representations and Practices* (Marburg: Metropolis Verlag, 2000), 143–73, 143.

²² Jack Goldstone, Jonathon Haughton, Karol Soltan, and Clifford Zunes, "Strategy Framework for the Assessment and Treatment of Fragile States," PPC IDEAS, Centre for Institutional Reform and the Informal Sector (IRIS), December 2003.

²³ *Ibid.*, 3.

²⁴ Vallings and Moreno-Torres, "Drivers of Fragility" (n. 5), 29.

understanding the causes of fragility will lead to aid programming which more effectively targets these causes.²⁵

It ought to be apparent already that fragility rests on a chain of concepts which are deeply contested and contestable. Terms such as legitimacy and effectiveness are controversial, and subject to a wide range of plausible definitions deriving from different normative and sociological frameworks; they are quintessentially theory-dependent concepts.²⁶ The serious difficulties of determining the effects of political institutions on political outcomes across very diverse social and historical realities are well-known.²⁷ Macrosocial questions such as what a state is or how one delimits the idea of society lack any consensus answers across disciplines and institutions, and the task of defining these terms even to posit them as objects of study is very difficult given their inseparability from history, belief, value, ideology, and political contestation. As Abrams noted in 1977, to reduce a state to its empirical functions or a mere aggregation of institutions is to overlook the way in which the state is always also an "ideological thing,"²⁸ which legitimates domination in part by successfully representing "political institutions in terms of cohesion, purpose, independence, common interest and morality without necessarily telling us anything about the actual nature, meaning or functions of political institutions."²⁹ Bourdieu makes a similar observation when he notes that one of the specific difficulties in studying the state "lies in the fact that . . . most of the writings devoted to the state more or less partake . . . of the *construction* of the state . . . From its inception, social science has been part and parcel of [the] work of the construction of the representation of the state which makes up part of the reality of the state itself."³⁰

The concept of the state found in the fragility literature is predicated on a highly particular image of the state, an ideal-type of the state form which is sometimes termed "Weberian." In this ideal-type, the state is a unified and coherent actor, set apart from other social organizations and sources of social power. Through its

²⁵ Carment et al., "Determinants of State Fragility and Implications for Aid Allocation" (n. 11), 2–3.

²⁶ See, e.g., the concise treatment of various arguments in Rodney Barker, *Political Legitimacy and the State* (Oxford: Oxford University Press, 1990). See also Christopher Thornhill, "Political Legitimacy: A Theoretical Approach Between Facts and Norms," *Constellations: An International Journal of Critical and Democratic Theory* 18 (2011), 135–69; for an overview of diverse normative and sociological approaches.

²⁷ See Alasdair MacIntyre, "Is a Science of Comparative Politics Possible?" in *Against the Self-Images of the Age: Essays on Ideology and Philosophy* (Schocken: New York, 1971), 260–79; Adam Przeworski, "Institutions Matter?," *Government and Opposition* 39 (2004), 527–40; Paul Pierson, *Politics in Time: History, Institutions and Social Analysis* (Princeton: Princeton University Press, 2004); Geoffrey Hawthorn, *Plausible Worlds: Possibility and Understanding in History and the Social Sciences* (Cambridge: Cambridge University Press, 1991).

²⁸ Philip Abrams, "Notes on the Difficulty of Studying the State," *Journal of Historical Sociology* 1 (1988): 58, 68.

²⁹ Ibid.

³⁰ Pierre Bourdieu, "Rethinking the State: Genesis and Structure of the Bureaucratic Field," *Sociological Theory* 12 (1994), 1, 3. See, in a similar vein, Migdal and Schlichte's distinction between "doing the state" and "seeing the state," and the regular divergence between practice and image: Joel Migdal and Klaus Schlichte, "Rethinking the State," in Joel Migdal and Klaus Schlichte (ed.), *The Dynamics of States: The Formation and Crises of State Domination* (Aldershot: Ashgate, 2005), 1–41.

officials, the state has its own preferences and can act on those preferences in order to change the behavior of others. The state enacts its domination through a uniform set of rules, backed by a credible threat of violence.³¹ At least for some Western European states, this ideal-type approximates an historical reality—although even among this relatively small group, variation is substantial throughout recent history.³² Importantly, for Weber, this ideal-type was not to be understood as a normative or normalizing typification of what qualifies as a “true” or “proper” state; rather, an ideal-type is an exaggerated abstraction from a particular historical reality, used only to better grasp the features of the phenomenon under analysis.³³ This ideal typical concept of the state—including the much-cited definition of it as an organization which exercises a monopoly of violence over a territory—was an abstraction that sought to identify a common form of political organization that emerged from a specific European experience. As Migdal and Schlichte observe, “an ideal-type is not itself a hypothesis, but it allows one to build hypotheses on deviations, variations and totally different forms. Once these differences are noticed, the need for a vocabulary of description and explanation becomes obvious.”³⁴ To the extent that the ideal type of the state is taken as the measure of state-ness, it becomes a norm rather than a heuristic for the development of explanatory or interpretive accounts of how a given political organization works (or does not work). The crypto-normative quality with which the Weberian ideal-type of the state has been imbued in the fragility discourse has meant that deviations from the type are taken as evidence of weaknesses or failures.

Fragility thus normalizes an idea of stateness in the sense described by Foucault: “Normalization consists first of all in positing a model, an optimal model that is constructed in terms of a certain result, and . . . [normalization then] . . . consists in trying to get people, movements, and actions to conform to this model, the normal being precisely that which can conform to this norm, and the abnormal that which is incapable of conforming to the norm . . . There is an originally prescriptive character of the norm and the determination and identification of the normal and abnormal becomes possible . . .”³⁵ Hacking points out that, in the history of statistical reasoning, normalization in this sense is also closely connected with—and

³¹ Weber in fact never completed a state theory. His concept of the state is scattered throughout his sociology of law and his omnibus work, *Economy and Society* (Berkeley: University of California Press, 1978).

³² See, for example, Thomas Ertman, *The Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe* (Cambridge: Cambridge University Press, 1997); and Samuel Finer, “State and Nation-Building in Europe: The Role of the Military,” in Charles Tilly (ed.), *The Formation of National States in Europe* (Princeton: Princeton University Press, 1975), 84–163.

³³ Max Weber, *Economy and Society, Volume One* (n. 31), 9, 20–1.

³⁴ Migdal and Schlichte, “Rethinking the State” (n. 30), 3.

³⁵ Foucault, *Security, Population, Territory* (n. 1), 57. Foucault describes this as disciplinary normalization or “normation” and contrasts it with statistical normalization, with the latter being a technique of 19th-century biopolitical governmentality. However, I cannot see why the two cannot in fact coexist with certain arrangements of power. For a critical analysis of how Foucault’s description of techniques of power defy easy periodization and may point to a “topological” method of studying forms of power, see Stephen Collier, “Topologies of Power: Foucault’s Analysis of Political Government beyond ‘Governmentality,’” *Theory, Culture and Society* 26:6 (2009): 78–108.

reinforced by—quantification. The plotting of data along normal distribution curves (a technique developed originally in metrology in order to approximate the exact position of celestial objects within a welter of differing measurements)³⁶ is applied to social phenomena to make inferences about normal qualities of populations, such as average height. Normal in this sense is not necessarily an ideal, but nonetheless serves as a standard against which deviations can be measured and thus a way of identifying the abnormal or pathological. The reality of the underlying quantity becomes a necessary assumption for the realism of the measure.³⁷ When applied to social and political objects, these techniques of quantification have fairly consistently led to a conflation of descriptive and evaluative notions of normal: a statistical regularity emerging from a set of observations transitions from an average to ideal property of a population, and becomes a standard to which deviant or significantly abnormal cases or groups *ought* to conform—perhaps through a range of reformist policy interventions.³⁸

The turn to quantification: Origins of USAID's index

The normalizing quality of the concept of fragility as it was articulated and invoked by development agencies would indeed be intensified through a turn to quantification. The interest in measuring fragility emerged more or less simultaneously with the coining of the concept itself, and a demand for “objective” means for applying the label was generated in part by the ambitions of proponents of the concept, and in part by resistance to the term from some constituencies.

The fragility concept focused attention on “state-ness” and state institutions not only among development agencies, but also in defense ministries and diplomatic corps. For a brief period, there was some interest in “whole of government” approaches to state fragility among key Western states, underscoring the idea that responding to security risks emanating from fragile states required integrated policies across security, development, and diplomacy actors.³⁹ Ambitions for an integrated approach to fragile states quickly foundered, however, with entrenched bureaucratic interests concerned to maintain control over their own domains of specialization. Patrick and Brown, discussing the US case, observe that the US approach is a “messy amalgam of the dominant preoccupations of the Department of Defense, State Department and USAID... The Pentagon views fragile and post-conflict states primarily through a national security prism...;

³⁶ Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990), 106; Desrosières, “How Real Are Statistics?”, (n. 13).

³⁷ Desrosières, “How Real Are Statistics?” (n. 13), 348; Hacking, *The Taming of Chance*, 108.

³⁸ See Desrosières, *The Politics of Large Numbers* (n. 4), 257–9; Hacking, *The Taming of Chance* (n.36), 119: “We obtain data about a governed class whose deportment is offensive, and then attempt to alter what we guess are relevant conditions of that class in order to change the laws of statistics that the class obeys. This is the essence of the style of government that in the United States is called liberal.”

³⁹ See Call, “The Fallacy of the ‘Failed State’” (n. 7), 1493; Patrick and Brown, “Greater than the Sum of Its Parts?” (n. 3).

the State Department is preoccupied with transforming a wider range of weak and war-torn states into effective democracies; and USAID regards state weakness as a developmental challenge to be addressed [by creating] the institutional foundations of good governance and economic growth.”⁴⁰ In the UK, an influential 2005 strategy document written in the Prime Minister’s office identified “effective states” as the answer to the problems of security, poverty, and governance in the developing world, and emphasized the importance of “mapping” risks of instability by developing “early warning” mechanisms which could identify and track risk factors driving conflict and fragility.⁴¹ But once again, the defense ministry, DFID, and Foreign and Commonwealth Office (FCO) did not implement common approaches or even definitions.

Nevertheless, all actors accepted in some form the common wisdom that weak or fragile states threatened national and global security, and ought to be identified and targeted with corrective policies. Specific departments and agencies began to explore methodologies for “early warning” of instability, although not necessarily in coordinated ways: the UK Prime Minister’s Strategy Unit proposed a “Countries at Risk of Instability” methodology, while the FCO considered statistical models of indicators for early warning.⁴² In the United States, the Central Intelligence Agency (CIA) is said to maintain a list of 50 “ungoverned areas” deemed to pose security risks, while the National Intelligence Council updates a political instability “watch list.”⁴³ In the early 1990s, the CIA established and funded the State Failure Task Force (later renamed the Political Instability Task Force) to explore methods of risk assessment and early warning systems for state failure.⁴⁴ The Polity IV dataset generated by this project, which spanned more than a decade, would become an important source for current fragility indexes, but it is reported that the Task Force also produced classified models and algorithms which are in use by the CIA to identify lists of countries at risk of instability.⁴⁵ The Canadian Defence agency, Defence Canada, has recently commissioned research into an Early Warning Model of state instability, including the use of indicators of state failure.⁴⁶

Among aid agencies working within the “fragile states” agenda, quantification developed as a response to operational needs, but also as a means of redeeming the promise that the fragility agenda could address “causes” of fragility. There was from early on a need for instruments to differentiate fragile states from non-fragile states, in order to identify states that are susceptible to the “fragile state strategy.” Perhaps the earliest instance of quantification connected with the notion of fragility was undertaken by the World Bank as part of its regular Country Policy and

⁴⁰ Patrick and Brown, “Greater Than the Sum of Its Parts?” (n. 3), 35.

⁴¹ Prime Minister’s Strategy Unit, “Investing in Prevention: An International Strategy to Manage Risks of Instability and Improve Crisis Response,” February 2005.

⁴² It seems that the FCO looked to the Carleton Indicators for Foreign Policy Index of State Failure as a basis for its framework, but classified the results of its analyses. See <<http://www4.carleton.ca/cifp/>>.

⁴³ Patrick and Brown, “Greater than the Sum of Its Parts?” (n. 3), 16–18, 36–7.

⁴⁴ Robert Adler, “The Crystal Ball of Chaos,” *Nature* 29 (2001): 480–1.

⁴⁵ Interview with USAID official, July 23, 2010.

⁴⁶ See Defence R&D Canada, “Indicators of State Failure, Phase II and Phase III Reports,” August 2010. The leaders of the Carleton project were contracted to produce the measures.

Institutional Assessment (CPIA), in which the Bank rates the institutional and policy arrangements of states receiving concessional loans through the Bank's International Development Association (IDA). The ratings are determined by Bank country staff and then vetted centrally, before a final number is agreed. Each state is evaluated based on 16 criteria, such as macroeconomic management and quality of public administration, and assigned a rating from 1–6 for each of these elements, and this is aggregated into an overall score. Countries are then ranked by score, and those falling below a certain score may become ineligible for further concessional loans due to a failure to improve their policy and institutional scores under the rating. The category of “fragile” state was introduced as a dispensation from this “performance-based” aid allocation: it was recognized that certain states heavily dependent on concessional loans could not meet policy and institutional performance standards because they had recently emerged from severe civil conflict or other economic and political upheaval.⁴⁷ Any state scoring less than 3.2 on the CPIA is deemed “fragile,” and may thus be eligible for continued access to IDA loans, and other targeted forms of assistance.⁴⁸ The CPIA rankings explicitly associate fragility with poor governance and economic vulnerability, but function largely as a means of “banding,” differentiating a special category of states in need of concessional lending, exemptions from performance criteria, and access to a special trust fund. They make no claims about risks of instability or conflict, and are expressly subjective in the sense that they are formed through the expert judgment of country staff, who assign a rating based on their assessment of various factors stipulated in explanation of the criteria.⁴⁹ Although it generates a number as an output, the process of fixing the number is a mixture of quantitative and qualitative assessments which is highly discursive.

Aid agencies shared the World Bank's interest in banding, but DFID and USAID—two agencies that had an early interest in quantification of fragility—expressed the ambitions of measurement more broadly. A DFID background paper on measuring fragility⁵⁰ explored measures that could be used to generate an index of “state capacity and willingness,” in order to identify fragile states and “inform decision-making” on policy responses to fragility. Developing indicators said to adequately proxy the “willingness and capacity” of a state was argued to help identify where DFID can “best effect progress” in institutional reforms through its policy interventions.⁵¹ A later document argued for the need to “anticipate fragility [and] avert it where possible,” in part by targeting institutional “drivers” of

⁴⁷ World Bank (International Development Association), “Operational Approaches and Financing in Fragile States” (n. 9), para. 7.

⁴⁸ *Ibid.*

⁴⁹ See World Bank, “Country Policy and Institutional Assessments: 2009 Assessment Questionnaire,” Operations Policy and Country Services, September 10, 2009.

⁵⁰ Michael Anderson, Andrew Branchflower, Magui Moreno Torres, and Marie Besancon, “Measuring Capacity and Willingness for Poverty Reduction in Fragile States,” PRDE Working Paper No. 6, DFID, January 2005.

⁵¹ *Ibid.*, 7.

fragility.⁵² USAID's background document entitled "Measuring Fragility"⁵³ contended that a quantitative measurement tool would "help USAID identify those countries at greatest risk for violent conflict and with the greatest need for early intervention... [and] also supply... data on the particular sources of fragility, and generally help measure country-level impact of USAID programs."⁵⁴ Thus, quantifying fragility for both of these agencies sought to identify not only classes of fragile states, but also to distinguish between them in terms of the *likely risk* of future instability, and to diagnose key causal determinants of fragility in a given state in order to target policy interventions and monitor change. Quantification had both predictive and diagnostic ambitions, engendering claims to knowledge about current and future states of the world.

DFID generated an unranked "proxy list" of 46 "fragile states" in its early policy statement on fragility,⁵⁵ but did not ultimately produce an index based on the background paper. It seems that FCO concerns about the political consequences of labeling states "fragile" applied a fortiori to a ranking of states by their degree of fragility: "the FCO is highly averse to using the term in their diplomatic discourse... FCO country desk officers are hesitant to agree to have their countries classified as fragile even for the purposes of an internal study."⁵⁶ It appears that DFID ultimately abandoned efforts at generating an Index, and relied instead on a "banding" approach to measuring fragility, taking the bottom two quintiles of the World Bank's CPIA rankings as identifying the set of fragile states.⁵⁷

At USAID, by contrast, devising an objective and transparent method of applying the label "fragile" was seen as essential to the authority of the concept, and the viability of a fragile state strategy. Part of the challenge was the lack of consensus concerning the meaning and definition of fragility itself. Proponents of quantification argued that a rigorous measurement methodology, which drew on the authority and prestige of academic experts, would underwrite the legitimacy of the concept itself, helping to generate consensus on which states ought to be labeled fragile. As one USAID official closely connected with the project stated, "We needed a diagnostic tool. We did not wish to rely on a [State Department or country] bureau's opinion, but wanted transparency, predictability and the absence of political interference."⁵⁸ A distrust of subjective judgment—and the expected reluctance of country experts in either the State Department or USAID to have "their" countries labeled fragile—also motivated the interest in numerical measures. To some degree, the objectivity and reality of fragility would be proven by its amenability to measurement and quantification.

⁵² Vallings and Moreno-Torres, "Drivers of Fragility" (n. 5), 5, 7.

⁵³ USAID, "Measuring Fragility: Indicators and Methods for Rating State Performance," June 2005.

⁵⁴ *Ibid.*, 1.

⁵⁵ See Annex 1, DFID, "Why We Need to Work More Effectively in Fragile States," 2005.

⁵⁶ Cammack et al., "Donors and the 'Fragile States' Agenda" (n. 3), 43.

⁵⁷ See Patrick Guillaumont and Sylviane Guillaumont Jeanneney, "State Fragility and Economic Vulnerability: What is Measured and Why?" European Development Report Background Paper, May 2009.

⁵⁸ USAID official, July 23, 2010.

USAID's interest in quantification thus seems to exemplify one aspect of the modern "trust in numbers" identified by Porter:⁵⁹ a distrust of esoteric and closely held subjective expertise (such as that of the country desk officer) in place of transparently calculative reasoning based on seemingly impersonal and unambiguous data-objects, which can be reviewed and argued about openly. But quantification here also instantiated another, related phenomenon, by which measurability becomes (implicitly or explicitly) a means of validating and establishing the quidity of a concept that is otherwise disputed. If something is countable and measurable—and is in fact regularly counted and measured—it's objective thing-ness seems more plausible, and its integration in decision-making and policy all the more credible. Desrosières gives the example of the concept of unemployment, which until the 1930s was not routinely measured due to wide divergences in how to define the phenomena and count it. As late as 1935, President Roosevelt questioned the reality of unemployment because nobody had "been able to define an unemployed person..."⁶⁰ Definitional controversies were overcome not by rationally motivated persuasion as much as by a—federally mandated—routinization of measurement by a prescribed instrument. Once unemployment was regularly and consistently measured by a designated bureaucracy, controversies over the reality of the *object* of unemployment became marginal even as arguments about *how to best measure it* persisted. Measurement and quantification, once embedded in networks of framing and acting, thus contribute to new "practical ontologies"⁶¹ in which new objects are solidified and become a common basis for action, coordination, and decision. By developing a method for measuring fragility, proponents of the fragility strategy at USAID sought to increase acceptance of the term itself.

Theorizing the state, prototyping strength, and counting what can be counted

USAID's definition of fragility emphasized "legitimacy" and "effectiveness" of the government of a state as the essential dimensions of its strength or fragility. Measuring fragility therefore required creators of the index to render countable qualities of a political order that are inherently unobservable and uncountable. Proxies had to be determined for each of these terms, based on theories about how a certain countable phenomenon (such as infant mortality or economic growth) was to be understood as representing a dimension of an uncountable property, such as legitimacy. Indeed, the very definitions of legitimacy and effectiveness required theoretical foundations, in order to articulate their claimed relationship with another non-observable term, "state strength" (or state weakness). As Munck and Verkuilen point out, specifying the meaning of a concept and its components "affects the entire process" of measurement, from data generation to aggregation,

⁵⁹ Theodore Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1990), chs 6–7.

⁶⁰ Quoted in Desrosières, *The Politics of Large Numbers* (n. 4), 205.

⁶¹ Bowker and Leigh Starr, *Sorting Things Out* (n. 16), Part 1.

and conceptualization is intimately linked with theoretical claims about the nature of the phenomenon one is seeking to measure.⁶²

In late 2003, USAID commissioned a background paper on fragility from the University of Maryland's IRIS Center (Center for Institutional Reform and the Informal Sector), which convened a series of roundtables of prominent academics to discuss approaches for defining, assessing, and "treating" fragile states. Key participants in the roundtables included Robert Bates, Daron Acemoglu, Jack Goldstone, Avner Greif, Timur Kuran, Thomas Schelling, and Barry Weingast. Drawn from either political science or economics, most of these academic experts are closely associated with a scientific literature that emphasizes the importance of institutional form and functioning, and regime type, on economic and political outcomes.⁶³ Bates and Goldstone were members of the State Failure Task Force, and Goldstone was a lead author of several of the Task Force's reports.⁶⁴ In the Strategy Paper written for USAID as a result of the deliberations of this group,⁶⁵ the strength or weakness of a state is defined in terms of the legitimacy and effectiveness of its *institutions*, especially those institutions charged with "the management of conflict." A strong state is one which successfully creates and maintains "the institutional mechanisms of 'neutral ground,'" which is composed on "those elements of the social life of the country that are *non-partisan* in ways relevant to the social conditions of that country, and which contain a range of *neutral legitimating factors such as legality, erudition and technical or professional competence*."⁶⁶ The "neutral ground" that characterizes state strength is thus *composed* of—identified with—the presence of institutions possessing certain qualities of effectiveness and legitimacy, which are in turn associated with features such as legality and technocratic competence.

These are clearly associative or correlative definitions, and causalities could run in either direction.⁶⁷ Gutierrez Sanin observes that definitions of state fragility and

⁶² G. Munck and J. Verkuilen, "Measuring Democracy: Evaluating Alternative Indices," *Comparative Political Studies* 35 (2002): 7.

⁶³ See, for example, Robert Bates, *When Things Fell Apart: State Failure in Late Century Africa* (New York: Cambridge University Press, 2008); Daron Acemoglu, S. Johnson and J. Robinson, "The Colonial Origins of Comparative Development," *American Economic Review* 91 (2001): 1369–401; Douglas North, Barry Weingast, and J. Wallis, *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History* (New York: Cambridge University Press, 2009); Avner Greif, *Institutions and the Path to the Modern Economy* (Cambridge: Cambridge University Press, 2006).

⁶⁴ Researchers associated with the Task Force continue to co-author academic journal articles deriving from the work done by the SFTF: see, for example, Jack Goldstone, Robert Bates, David Epstein, Ted Gurr, Michael Lustik, Monty Marshall, Jay Ulfelder, and Mark Woodward, "A Global Model for Forecasting Instability," *American Journal of Political Science* 54 (2010): 190–208.

⁶⁵ Goldstone et al., "Strategy Framework" (n. 22).

⁶⁶ Jack Goldstone, Jonathon Haughton, Karol Soltan, and Clifford Zinnes, "Strategy Framework for the Assessment and Treatment of Fragile States," Executive Summary, June 2004, 5–6 (my emphasis).

⁶⁷ *Ibid.*, 8. Additional glosses on the meaning of effectiveness and legitimacy are provided in the document. Effectiveness is defined as "the degree to which a state has the administrative capability and resources to carry out the tasks of governance," where governance includes "a disciplined military and bureaucracy" and "intelligence and administrative capability. Legitimacy is further described as "rulers being judged—by both elites and popular groups—as being reasonably fair and just in their exercise of power."

	Effectiveness	Legitimacy
Political		
Economic		
Social		
Security		

Fig. 6.1 PESS-EL Matrix for State Assessment

Source: Goldstone et al., "Strategy Framework for the Assessment and Treatment of Fragile States: Executive Summary," (USAID and the IRIS Center), 9.

state strength used in existing Indexes are prototypical, in the sense that they identify typical instantiations of strength or failure and derive common characteristics.⁶⁸ The strategy paper and subsequent Index developed by USAID share this feature with other Indexes, taking qualities observed in typical non-fragile states and identifying them as the essence of strength. The fundamental difficulty with prototyping is that "the characteristic traits of the phenomenon are collapsed with putative causes and consequences . . . This is a bit like defining cancer as a consisting of smoking, uncontrolled growth of cells, and family crisis."⁶⁹ Such prototypical definitions do not specify causal pathways between observed characteristics and fragility or strength, and provide no means to understand the relative role of each characteristic or how they interact with each other temporally, and do not propose how each dimension can be logically assembled into a single notion of strength or weakness.⁷⁰ Prototyped definitions also intensify the normalizing effect of the concept: to be a strong state is to be *like* x, y, and *not like* b, c.

In their attempt to "operationalize" the concepts of "effectiveness" and "legitimacy" in order to render them useful for the assessment of state fragility, the strategy paper authors develop a four-row by two-column accounting matrix (Figure 6.1), in which rows represent "four dimensions of state-society relations: political . . . ; economic . . . ; social . . . ; and security," while the columns represent effectiveness and legitimacy for each of these dimensions. An assessment of a state's fragility entails gathering evidence of the effectiveness and legitimacy of the state's political, economic, social, and security functioning, and evaluating this evidence against a benchmark of state strength:

States that cannot enforce laws, collect sufficient revenue, or administer and control their territory are *politically ineffective* . . . For the *economic* dimension . . . [w]e ask if the state can provide food security, access to water, land, jobs, stable currency, and other resources essential to maintaining material life. . . . Legitimacy of the dimensions of state capacity can be made operational in terms of 'fair shares'. Politically, if specific elite or popular groups are systematically or explicitly excluded from power, there is likely to be a perception of low legitimacy along the political dimension . . . For security, if the security apparatus works to favor or repress a specific social group, works erratically or indiscriminately against the

⁶⁸ Francisco Gutierrez Sanin, "Evaluating State Performance: A Critical View of State Failure and Fragility Indexes," *European Journal of Development Research* 23 (2011): 24.

⁶⁹ Ibid.

⁷⁰ The final point leads to a problem of aggregation, which will be discussed further below.

population at large, or fails to predictably and adequately enforce property rights then there is likely to be a perception of low legitimacy along this dimension.⁷¹

This decomposition particularizes further the characteristics said to be associated with effectiveness and legitimacy. It also interpolates some claims about causal relationships between certain dynamics and the extent of effectiveness and legitimacy (e.g., exclusion “is likely” to diminish legitimacy). These causal claims are theoretical, abstracted from unspecified historical experiences but not rigorously demonstrated at a conceptual or empirical level in the text. As such, the claimed relationships between observable trends (such as non-enforcement of property rights) and unobservable qualities, such as legitimacy, are not established; rather, they are stipulated, as part of the typological construction of fragility.

The accounting matrix, with its four-by-two enumeration of dimensions of effectiveness and legitimacy, would become a template for developing measures of fragility that would serve as inputs into the Index. A way had to be found to quantify each of eight possible dimensions of effectiveness and legitimacy,⁷² rendering them *countable* and so susceptible to aggregation in a one-dimensional rank ordering. Countability requires that data be found or created which successfully measure the phenomena sought to be counted. Where the phenomena (however well specified) are unobservable, it is necessary to resort to proxies which can be observed and counted, implying a theory of the relationship between the proxy and the attributes of the underlying concept. Counting is always reasoning. At this stage of the design of the USAID Index, the analytical and empirical difficulties encountered at the stage of defining state strength and fragility are reincarnated at a greater level of particularity.

Analytically, concepts such as political legitimacy or social legitimacy are not well specified in the sense that their stipulated attributes are themselves unobservable. For example, political legitimacy is defined as “political institutions and processes that are transparent, respect societal values, and do not favor particular groups.”⁷³ Each of these terms is non-self-evident, and what might qualify as a measure of each of these attributes is unclear, requiring further disaggregation and conceptualization. Moreover, each of the attributes of political legitimacy is difficult to logically distinguish from attributes of other dimensions of legitimacy, such as social legitimacy or security legitimacy. The former is defined as “tolerance for diversity, including opportunities for groups to practices . . . cultures and beliefs,” while the latter is defined as “military and police services that are provided equitably and without violation of civil rights.” The problem with logically non-exclusive definitions of attributes of unobservable variables such as political legitimacy, is *redundancy* and *conflation*.⁷⁴ How do we know that a measure of tolerance is not also a measure of respect for societal values? If we do not know, how can we avoid

⁷¹ Goldstone et al., “Strategy Framework for the Assessment and Treatment of Fragile States,” Executive Summary (n. 66), 10 (emphasis original).

⁷² USAID, “Measuring Fragility” (n. 53).

⁷³ Ibid., 2.

⁷⁴ See Munck and Verkuilen, “Measuring Democracy: Evaluating Alternative Indices” (n. 62), 12.

problems of double counting, or inaccurate differentiation between two measures of the same underlying phenomena? The analytical difficulties of conceptualization in fact point to the deep difficulty of measurement of multi-dimensional characteristics such as legitimacy, a difficulty that is at once compounded and made less visible where measures are aggregated into a one-dimensional number, as in an index.

Conceptual problems of this kind are aggravated in the attempt to identify proxies which can be used as observable data representing unobservable phenomena. Given that definitional problems remain severe, it is a puzzle how proxies could be determined: how could one differentiate between good and bad proxies for a concept, when the concept itself is not well specified? The creators of the USAID Index sidestepped this puzzle by *substituting the proxy for the concept*. Instead of identifying measures that successfully capture attributes of the concepts sought to be measured, the creators identified pre-existing indicators which they contended measured political, social, economic, and security *outcomes or perceptions*, and maintained that these outcomes or perceptions *amounted to* measurements of the eight dimensions of effectiveness and legitimacy decomposed by the accounting matrix.

Definitions of effectiveness and legitimacy developed by the strategy paper were thus decoupled from the identification of measures, with an emphasis instead on where “good data” could be found that could conceivably be related to the measurement of effectiveness and legitimacy. The result is a bricolage of datasets chosen because of their coverage, frequency of updating, assumed credibility of their source, and some notion of robustness. But the critical question of how and why the outcome or perception data is in fact a measure of the unobservable variable is answered only with the introduction of further, diverse, theoretical, and factual claims. For example, among the proxies chosen to measure “political effectiveness” is the government effectiveness indicator generated by the World Bank’s Governance Matters dataset, based on an aggregation of surveys of *perceptions* of the quality of public service. The rationale for choosing this as a component measure of political effectiveness is that “the quality of public service provision is a good, directly observable outcome of effective governments.”⁷⁵ Yet, the quality of public service provision is not at all directly observable, which is why the Governance Matters dataset relies on subjective perception surveys to develop a measure.⁷⁶ The relationship between perception and reality is taken as linear and direct by the USAID Index, whereas it has been commonly observed that it is difficult to know what underlying quality is being measured by such surveys: “Are they really making judgments about the quality of governance and particular institutions, or are they implicitly answering the question, ‘how do you think things are going today in country X (perhaps compared implicitly to countries in the region)?’”⁷⁷

⁷⁵ USAID, “Measuring Fragility” (n. 53), 6.

⁷⁶ See Fearon, “Do Governance Indicators Predict Anything?” (n. 8), 27.

⁷⁷ Ibid., 28. See also Andrew Williams and Abu Siddique, “The Use (and Abuse) of Governance Indicators in Economics: A Review,” *Economics of Governance* 9 (2008): 131–75; Svend-Erik Skaaning, “Measuring the Rule of Law,” *Political Research Quarterly* 63 (2010): 449–60; Katharina Pistor, “Re-construction of Private Indicators for Public Purposes,” this volume, pp. 165–79.

Because the concept of “political effectiveness” remains underspecified for the USAID Index, we cannot begin to determine how this indicator *could* be theorized as a component measure of this dimension or what weight to give it.

In numerous other choices of indicators as component measures, a theory of cause and effect is asserted between the outcome measure, and effectiveness or legitimacy. But the relationship is not demonstrated and is neither self-evident nor intuitively obvious. Thus, data from the Minorities at Risk (MAR) dataset is used to derive an indicator for the “percentage of population experiencing political discrimination,” which is asserted to measure an aspect of political legitimacy: “The ability for all groups to participate effectively in the political process is at the core of the concept of political legitimacy. If groups feel excluded or marginalized in the political process, they are likely to seek non-political⁷⁸ means to pursue their interests.”⁷⁹ The MAR in fact counts discriminatory *policies* rather than discriminatory *outcomes* (such as the feelings or conduct of minorities), and the relationship between policies and outcomes cannot be assumed to be linear (as we know from the ongoing controversy concerning the redistributive effects of affirmative action). But equally unclear is how discriminatory policies necessarily diminish “political legitimacy,” in the absence of a large variety of other intervening causes: conflict, social mobilization, external criticism and sanction, foreign assistance to would-be rebel groups, to name but a few possibilities. Only by adopting a strictly normative idea of legitimacy (discrimination violates the principle of equal treatment, which is a necessary part of legitimacy) could we say that discriminatory policy is of itself an indicator of the legitimacy of a political order. It is equally conceivable, however, that the selective and strategic persecution of a minority might *enhance* the legitimacy of a political order in the eyes of its population, who are encouraged to blame their woes and anxieties on this minority and regard them as a threat against which they must be protected by the state. Such discrimination would enhance what Weber termed the sociological legitimacy of the state.

The USAID Index identified 33 outcome indicators claimed to measure effectiveness and legitimacy across political, security, economic, and social dimensions (Figure 6.2). Each of these indicators suffers from one of the two difficulties described above: it is a subjective perception index, with an erroneous assumption that perception stands in a linear relationship to reality, rather than at best a very “noisy” approximation of complex trends; or it equates a policy measure with an outcome, or assumes that outcome data reflects effective policy.⁸⁰ But in all cases, the relationship between the indicator and higher-order variable of effectiveness or legitimacy is dubious, both because the concepts of effectiveness and legitimacy are poorly specified, and because the arguments made for why the indicator adequately

⁷⁸ It is unclear what is meant by this, but it seems that non-political is equated with violent.

⁷⁹ USAID, “Measuring Fragility” (n. 53), 10.

⁸⁰ The latter assumption may be more plausible in the case of infant mortality or literacy, which the Index counts as a measure of “social effectiveness” of a state: *see*, for discussion, Gary King and Langche Zeng, “Improving Forecasts of State Failure,” *World Politics* 53 (2001): 652.

Outcome Indicators

	Effectiveness	Legitimacy
Political	Quality of public service (World Bank Governance Indicators)	Nature of political participation (factionalism) (Polity IV)
	Number of coups d'etat in last five years (Polity IV)	% of population experiencing political discrimination (Minorities at Risk)
	Government revenues as % of GDP (IMF, World Bank, CIA factbook)	Extent of political participation in selecting government (World Bank Governance Indicators)
		Asylum requests as % of population (UN High Commissioner for Refugees)
Security	Intensity of most severe armed conflict (Polity IV)	State use of political terror (Political Terror Scale)
	Size of displaced population (US Committee for Refugees)	Extent of state repression of citizens (PITF/Polity IV)
	Proportion of area affected by ethnic or revolutionary war (Polity IV/PITF)	Presence/change in support for militant groups (Minorities at Risk)
Economic	Three-year change in real GDP (PPP) per capita (World Bank Development Indicators) (WBDI)	% of population experiencing economic discrimination (Minorities at Risk)
	Change in foreign investment (WBDI)	Corruption (World Bank Governance Indicators)
	Poverty rate (% of population living on <\$2 [PPP]/day (WBDI)	Extent of rule of law/protection of property rights (Index of Economic Freedom, Heritage Foundation)
	Primary commodity exports/total exports (WBDI)	Number of days to start a business (WBDI)
	Three-year inflation rate (WBDI)	
Social	Infant mortality rate (WBDI)	Male/female literacy ratio (WBDI)
	Youth literacy rate (WBDI)	Male/female life expectancy ratio (WBDI)
	Change in % of population living with AIDS (UNAIDS)	% of GDP spent on military (SIPRI)
	DPT and measles immunization rate (WBDI)	Deviance from GDP-predicted infant mortality (WBDI)
	% of population with access to improved water supplies/sanitation (WBDI/WHO)	Deviance from GDP-predicted primary school completion rate (WBDI)
		Cultural and religious freedoms (WBDI)

Fig. 6.2 Proposed Outcome Indicators

Source: USAID, "Measuring Fragility: Indicators and Methods for Rating Star Performance," June 2005, 3.

measures these terms are unsubstantiated. The result is a collection of indicators drawn from a large variety of datasets with widely disparate methodologies, which *stand in for*, rather than measure, concepts of effectiveness and legitimacy. The numbers *become* the concepts rather than represent or approximate them, creating a misleading impression that analytical problems of specification, commensuration, and aggregation have been resolved because all inputs now appear as equivalent quantities.

Absorbing uncertainty, commensurating, aggregating

The reliance on pre-existing datasets to generate numbers claimed to measure the effectiveness and legitimacy of a state presupposes a vast information infrastructure,

in which numerous sets of concepts have already been defined, measurements devised, data gathered, and aggregation questions addressed. The datasets and indicators used by the USAID Index derive from a broad range of sources, institutions, and problematics: many are concerned with measuring governance and regime type,⁸¹ while others report economic and epidemiological data.⁸²

Every kind of quantification confronts challenges to its validity and reliability, even apparently “well-settled” economic data such as GDP. The historicity of statistical reason shows that the “style of reasoning” that characterizes modern statistics is sufficiently rich and deep to generate a large matrix of arguments that can be used to justify and contest the truth-value of any given statistical fact.⁸³ What succeeds in becoming an established or widely used measure—a settled statistical fact, “hard” rather than “soft” data—may reflect less its strict determination by authoritative methods, than its enrolment in institutions and networks that regularize and defend its use as coherent with canons of statistical reasoning.⁸⁴ The uncertainties entailed at each stage of the creation of a statistical fact, from conceptualization to data generation to inferences from data, are to some degree absorbed and erased once the fact is “in use.”⁸⁵

Of the datasets relied upon by USAID to generate index inputs, some are regarded as well-settled, already widely in use and “hard” enough to stand in as real measurements of the phenomena they purport to measure, and not something else. Others remain much more controversial, and the relationship between measure and referent is regularly contested. By summarizing data from these datasets into a one-number-per-country manner, the USAID Index works to absorb and erase the uncertainty surrounding the data and what it measures. This is perhaps most clear in the reliance on datasets that purport to measure the quality of governance, extent of discrimination, human rights abuse, and regime type. The governance indicators are already aggregations of a large number of subjective perception surveys, and as noted above, it is difficult to determine whether each underlying survey measures institutional quality or some more vague set of perceptions.⁸⁶ The Index also uses indicators derived from the Minorities at Risk, Polity

⁸¹ Examples of such datasets used in the USAID Index include Governance Matters (World Bank Institute); Minorities at Risk (University of Maryland); Polity IV (Political Instability Task Force); the Political Terror Scale (compiled by M. Gibney); Index of Economic Freedom (Heritage Foundation); Percentage of GDP used on Military Spending (SIPRI).

⁸² From the World Bank’s World Development Indicators, the following indicators are used: Three-year change in Real GDP per capita, Change in Foreign Investment, Poverty Rate, Primary Commodity Exports as a proportion of Total Exports, Inflation, Infant Mortality, Deviance from GDP-expected mortality, Youth Literacy, DPT and Measles Immunization Rates, Access to Improved Water, Male/Female Literacy and Life Expectancy Ratios. A Change in the Percentage of the Population Living with HIV/AIDS indicator is drawn from UNAIDS.

⁸³ Ian Hacking, “Statistical Language, Statistical Truth and Statistical Reason: The Self-Authentication of a Style of Scientific Reasoning,” in Ernan McMullin (ed.), *The Social Dimensions of Science* (South Bend, IN: Notre Dame University, 1992).

⁸⁴ See Desrosières, *The Politics of Large Numbers* (n. 4), chs 3, 4.

⁸⁵ Wendy Espeland and Mitchell Stevens, “A Sociology of Quantification,” *Archive of European Sociology* 49 (2008): 421–2.

⁸⁶ Fearon, “Do Governance Indicators Predict Anything?” (n. 8), 28.

IV, and Political Terror Scale to measure discrimination, regime type, and human rights abuses, respectively. Like the USAID Index into which they are aggregated, each of these datasets attempts to count inherently uncountable, qualitative phenomena. In order to do so, each relies on elaborate coding processes whereby a coder allocates a number (0 or 1, 0–2, 0–10, 0–5, depending on the dataset) based on his or her interpretation of qualitative information (“content analysis”) about the political and social circumstances in the relevant state, or based on whether a given event has taken place (e.g., a military coup) in a given year. Vast swathes of history and politics across numerous, heterogeneous, political orders are thus commensurated and transformed into a number *rating* of the country situation. Consistency in coding and assignment of rating numbers is sought through detailed specification of coding rules, training of coders, and assignment of more than one coder to the same countries in order to test “inter-coder reliability.”⁸⁷ Transparency is maintained by the publication of code books, and also of information concerning the extent of discrepancy between different coders’ rating of the same country situation. “Reliability” of such ratings is clearly a benchmark or best practice—consistency with what contemporary statistical reason considers adequate method: “a reliable indicator is one whereupon different people apply an identical measurement procedure to identical information and come up with identical scores.”⁸⁸

Reliability, in the sense of complete identity between information and ratings across different coders, is never achieved. But even where very high levels of consistency (above 90 percent) are achieved, fundamental uncertainties remain as to what a given rating actually represents in terms of a state’s political and social reality. One obstacle is endogeneity, as described by Fearon in relation to datasets coding discrimination: “these measures of political exclusion and discrimination are based on the subjective judgments of diverse coders, trying to code somewhat impressionistic things. Countries where there has been no ethnic conflict and where ethnic relations have been calm are for that reason judged to have [little] exclusion . . . one can reasonably worry that a coder’s knowledge that there was an ethnic conflict in a country increases the probability that he or she judges that . . . groups were discriminated against or politically excluded.”⁸⁹ McCormick and Mitchell note that even where a high level of reliability is achieved in coding human rights violations, there is a basic problem of the *meaning* of a given rating: “human rights violations differ in type not just amount, such that they cannot be clearly represented on a single scale. Imprisonment and torture are different types of government activity [involving] differing uses of governmental resources and capabilities, and differing costs for the government . . . Regimes use different mixes of

⁸⁷ A description of best practices of coding and verification can be found in David Cingranelli, “The Cingranelli and Richards (CIRI) Human Rights Data Project,” *Human Rights Quarterly* 32 (2010). The CIRI dataset is not used by the USAID Index, but the process of coding described by Cingranelli applies more or less directly to the datasets used by the Index.

⁸⁸ *Ibid.*, 409.

⁸⁹ Fearon, “Do Governance Indicators Predict Anything?” (n. 8), 16.

methods of political control, a variation missed by a one-dimensional scale.”⁹⁰ Cingranelli, one of the authors of a major human rights scale, concludes that, due to inadequacies in the consistency of underlying data quality and availability, a human rights scale is not reliable beyond three or four categories, and ought not to be used to derive an ordinal ranking.⁹¹

The key point here is not that these various attempts to develop cross-national datasets of political qualia are necessarily meaningless; rather, it is that we cannot readily reach conclusions about what they mean, or what a higher or lower number within a dataset represents in terms of the “effectiveness and legitimacy” of a state. Yet global scale indexes such as the USAID Index perform a further *aggregation of aggregations*: the interpretive ambiguity, noisiness, uncertainty, and measurement error which are an inherent part of each of these datasets are simply erased through addition into a single score. Espeland and Stevens note March and Simon’s description of the way in which uncertainty is absorbed as information travels upward within an organization:

“Raw” information typically is collected and compiled by workers near the bottom of organizational hierarchies; but as it is manipulated, parsed and moved upward, it is transformed so as to make it accessible and amenable for those near the top, who make the big decisions. This “editing” removes assumptions, discretion and ambiguity, a process that results in “uncertainty absorption”: information appears more robust than it actually is. As March and Simon put it: “Uncertainty absorption takes place when inferences are drawn from a body of evidence, and the inferences instead of the evidence itself, are then communicated.”⁹²

This insight has particular application to the new generation of global scale indexes. What is moved upward to a higher level of abstraction and aggregation is not “raw” information, but already highly artifactual composites and aggregates, the result of laborious construction, parsing, concept-making, judgment-calling, and contestation. Transmitted upwards are sets of inferences and claims bundled into a one-dimensional figure, which the higher-level index then bundles further.

In her work on commensuration,⁹³ Espeland highlights the way in which quantification and ranking rest on commensuration. Qualitative distinctions are transformed into quantitative ones, and “difference is expressed as magnitude

⁹⁰ James M. McCormick and Neil Mitchell, “Human Rights Violations, Umbrella Concepts and Empirical Analysis,” *World Politics* 49 (1997): 513–14. In relation to the Polity IV dataset, Gleditsch and Ward note, “Categorical data encompass a small number of subdimensions that interact in non-obvious ways . . . Data are multipath—there is a wide variety of ways in which polities can receive a single scale value. Vastly different temporal, spatial and social contexts support the same democracy and autocracy scale values.” Kristian S. Gleditsch and Michael D. Ward, “Double Take: A Reexamination of Democracy and Autocracy in Modern Polities,” *Journal of Conflict Resolution* 41 (1997): 380.

⁹¹ Cingranelli, “The Cingranelli and Richards (CIRI) Human Rights Data Project,” (n. 87), 406, 408. The Political Terror Scale, used by the USAID Index, has five categories and purports to measure human rights “conditions” rather than the narrower “practices.”

⁹² Espeland and Stevens, “A Sociology of Quantification” (n. 85), 422. Online reference omitted.

⁹³ Wendy Espeland, “Commensuration and Cognition,” in Karen Cerulo (ed.), *Culture in Mind: Toward a Sociology of Culture and Cognition* (London and New York: Routledge, 2002), 63–88.

according to a shared metric.”⁹⁴ Disparate objects, qualities, entities, are placed in relation to each other in new ways, unified over an apparently common category and internally differentiated within that classification. In the case of the USAID Index, an expansive architecture of commensuration—represented by the datasets from which each of the 33 indicators is sourced—is commensurated further. The result is a new quotient: stateness. The state, which is simultaneously material and ideal, practice and concept, becomes a quantity with its own numerical properties. Numbers enable rankings, allowing for the expression of a hierarchy or spectrum of stateness.

USAID’s Index ranking for each country is obtained by averaging the “effectiveness score” and “legitimacy score” for each country into a “fragility score.” The effectiveness and legitimacy scores are the aggregations of component indicators selected as proxies for each of the four dimensions of effectiveness and legitimacy. The step of aggregation is indispensable for deriving a single score for each state, allowing a seemingly transitive rank ordering to be obtained. Rank ordering is one of the most seductive features of unidimensional indexes: it creates an impression of precise differentiation between objects (in this case states) in respect of a complex property (stateness) that seems to have been successfully reduced to a unidimensional measure. As Merry, Kingsbury, and Davis note, indicators such as the Human Development Index gain authority in part because they become a “short-hand” for a country’s circumstances and rankings imply standards to be met or shortcomings made legible (and perhaps actionable).⁹⁵ Rank ordering was also an essential part of USAID’s intended use of the Index: ordinal ranking would facilitate its diagnostic use by helping identify which countries ought to be given priority in the roll out of the Fragile States Strategy, because of their high fragility score (using components of the fragility score to determine what the drivers of fragility are), and rankings supported a predictive use of the Index, by allowing identification of which states’ fragility scores were deteriorating over time, and so were “at risk” of instability.

For rank ordering to fulfill these ambitions of diagnosis and prediction, the differences between country ranks must in fact capture some real distinctions and interactions in the dimension of fragility—and these differences must be amenable to transitive ordering in the sense that if Sierra Leone is ranked above Cambodia, and Cambodia is ranked above Cameroon, it must be true that Sierra Leone is more fragile than Cameroon.⁹⁶ For fragility and its underlying dimensions (and the indicators of those dimensions) to be susceptible to this treatment, the problems of conceptual specification and proxy identification pointed out earlier must have been suitably resolved:

⁹⁴ Wendy Espeland, “Commensuration and Cognition” (n. 93), 64.

⁹⁵ See generally Kevin Davis, Benedict Kingsbury, and Sally Engle Merry, “Global Governance by Indicators,” this volume, pp. 3–28.

⁹⁶ A correct aggregation function is a set of rules that attributes to a vector of characteristics a single element in the range and additionally: (a) behaves monotonically; and (b) respects boundary conditions: G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners* (New York: Springer Berlin Heidelberg, 2007) cited in Gutierrez Sanin, “Evaluating State Performance” (n. 68), 32.

First, the analyst must make explicit the theory concerning the relationship between attributes [dimensions]. Second, the analyst must ensure that there is a correspondence between this theory and the selected aggregation rule, that is, that the aggregation rule is actually the equivalent formal expression of the posited relationship. For example, if the aggregation of two attributes is at issue and one's theory indicates that they both have the same weight, one would simply add the scores of both attributes. If one's theory indicates that both attributes are necessary features, one could multiply both scores, and if one's theory indicates that both attributes are sufficient features, one could take the score of the highest attribute. In this regard, then, it is crucial that researchers be sensitive to the multitude of ways in which attributes might be linked and avoid the tendency to limit themselves by adherence to defaults, such as additivity.⁹⁷

To put it another way, an aggregation function for a multidimensional phenomenon like "fragility" requires the solving of two kinds of problem: determining direction of causalities between fragility and its dimensions, and between dimensions,⁹⁸ and determining a *numeraire* which allows one to calculate how many units of dimension A substitute one unit of dimension B. Without resolving these two puzzles, rank ordering beyond pair-wise comparison is arbitrary.⁹⁹

USAID's background paper on "Measuring Fragility" contains no discussion of the difficulties of aggregation, and the Index states that Principal Component Analysis is used to derive a weighted average for the component indicators for effectiveness and legitimacy. The average is then standardized to make the mean for all countries 0, and a country's deviation from the mean becomes its fragility score; a positive score connotes higher fragility; a negative score reflects lesser fragility. This generates the appearance of unidimensionality, but only by placing the statistical cart before the theoretical horse, sidestepping the validity problems raised by poor specification, doubtful proxies, and no theory of aggregation.¹⁰⁰

A number tells a thousand words (and pictures): Formalizing practices, panoramas, ontologies of association

Within the style of reasoning of contemporary statistical method, the numbers generated by USAID's Index do not appear robust or easily defensible. They are not very good numbers. Rather than consider how one might more successfully measure stateness, it seems to me that a pressing puzzle is the existence and growth

⁹⁷ Munck and Verkuilen, "Measuring Democracy: Evaluating Alternative Indices" (n. 62), 24.

⁹⁸ Recall that the USAID Index asserts 8 dimensions of fragility (4x2). King and Langche note that establishing interactions for a six-dimensional space (assuming, plausibly linear and non-linear interactions) is "almost incomprehensibly immense." King and Langche, "Improving Forecasts of State Failure" (n. 80), 639.

⁹⁹ See Gutierrez Sanin, "Evaluating State Performance" (n. 68), 30–3 for detailed discussion.

¹⁰⁰ Gutierrez Sanin rightly points out that for indexes of state fragility, "the aggregation function is a substantial part of the theory" of the fragility concept itself.

of demand for rankings based on relatively poor concepts, data, and methods.¹⁰¹ As Lampland suggests, it may well be a mistake to assume that the effective uses of a number depends upon their *veracity* or some defensible argument for veracity.¹⁰² Quantification is a formalizing practice that can serve a variety of purposes, even if a number's claim to objectivity and scientificity cannot be readily cashed out under the prevailing exchange rate for scientific method: "Numbers are instruments, not simply transparent signs."¹⁰³ But what are they instruments for, in this case?

It is difficult to answer without more extensive fieldwork, but we can note that the initial proposed operational use for the USAID Index—as a means of selecting countries targeted for the roll out of the Fragile State Strategy—only occurred once. Four states in which the Fragile State Strategy was to be piloted were chosen using the Index in 2005–6, and USAID sought discretionary funding from Congress to implement specifically crafted aid programming in these states. However, by mid-2006, Secretary of State Condoleezza Rice declared her "transformational diplomacy" initiative, which transferred much decision-making concerning aid programming to the State Department's "F" Bureau. The bureau did not embrace the Fragile State Strategy, and the use of the Index as a selection and diagnostic tool did not continue.¹⁰⁴ The pilot programming was terminated after one of the selected states, Ethiopia, learned of its classification as "fragile" and objected to the US Ambassador.¹⁰⁵ The possibility that the ranking of countries on the Index could have direct consequences for aid allocations did not transpire.

Strikingly, however, the generation of the Index itself was not discontinued, and it was produced annually from 2006. The extent of its dissemination is unclear, but it is said that "anyone with a dot gov email address can receive it" and requests to be added to the distribution list have been steadily growing. What each end-user does with the information must be explored further to fully account for how the Index impacts on decision-making, if at all. As March and Feldman note, many bureaucracies accumulate information without using it as a basis for decision, but in order to send a signal that they have "taken it into account."¹⁰⁶ Within USAID, the Index is used as an "early warning tool, to focus attention of bureaus on specific countries at risk. It condenses a lot of information . . . [and] . . . the rank ordering can be provocative . . . We can use it as a basis for starting a conversation."¹⁰⁷

¹⁰¹ See Christopher Hood, Ruth Dixon, and Craig Beeston, "Rating the Rankings: Assessing International Rankings of Public Service Performance," *International Public Management Journal* 11 (2008): 298–398 for similar puzzlement in relation a range of rankings.

¹⁰² Martha Lampland, "False Numbers as Formalizing Practices," *Social Studies of Science* 40 (2010): 377–404.

¹⁰³ *Ibid.*, 383.

¹⁰⁴ Dane F. Smith Jr, *U.S. Peacefare: Organizing American Peace-Building Operations* (Santa Barbara: Praeger, 2010), 136.

¹⁰⁵ USAID Official, July 23, 2010.

¹⁰⁶ Martha S. Feldman and James March, "Information in Organizations as Signal and Symbol," *Administrative Science Quarterly* 26 (1981): 171–86. My thanks to Bruce Carruthers for directing me to this reference.

¹⁰⁷ USAID Official, July 23, 2010.

Synoptic, attention-grabbing, provocative: these terms are used to describe the effects of ordinally ranking state-strength, even as they acknowledge that the methodological bases for ranking may not be persuasive to many.¹⁰⁸ The reduction of a complex, multidimensional reality (nothing less than the political and social order of a territory) to a single number is admitted by most to be susceptible to criticism. But as argued above,¹⁰⁹ the category of fragility seems to demand a means of classifying and comparing across vast scales of people, space, and history. Closely intertwined with policy demands to globally forecast security risks and target sources of disorder within states, it is a problematic and heuristic which elicits panoramic techniques—a tendency evidenced by the numerous other attempts to quantify and index state fragility.¹¹⁰ Panoramas, as Latour argues, are artifacts that aspire to see the whole and *show* it in an ordered and coherent way: “What is so powerful in those contraptions is that they nicely solve the question of staging the totality, of ordering the ups and downs, of nesting ‘micro,’ ‘meso’ and ‘macro’ into one another . . . They collect, they frame, they rank, they order, they organize.”¹¹¹

A panorama of stateness, such as that presented by the USAID Index, integrates a theory of political order (founded in institutions, achieved through technical efficiency and legitimate conflict resolution) with a raft of measures and numbers to generate a macro-level snapshot of degrees of effectiveness and legitimacy of 161 states across the world. Rendering degrees of political order calculable in this manner also changes the way in which it can be made visible and represented: unlike an expert opinion, a human intelligence report, or a historical monograph, the judgments and narratives embedded in a unidimensional number are easy to transpose into two-dimensional space, whether tabular (a list color coded by degrees of fragility) or pictorial (a map color coded by degrees of fragility or instability risk). In the color code used by USAID, most fragile is represented by red, high fragility by orange, moderate fragility by yellow, “some” fragility by green, and low fragility by blue. The map creates a regionalized picture of fragility, in which there are zones of the world suffering lesser or greater degrees. Unsurprisingly, those zones in which red, orange, and yellow predominate overlap significantly with underdeveloped or lesser developed countries. Those zones where green and blue predominate are first world/Northern countries and regions, with some exceptions. The effect of this aesthetic is striking: zones associated with poor governance and instability are marked with a color connoting danger/threat/heat (red, orange), and zones associated with good governance and stability are marked with a color connoting safety/ease/calm or cool (green, blue). Indeed, the aesthetics of the fragility map resemble a “heat map,” sometimes used to represent risks of contagion and spillover of epidemics or (more recently) systemic financial crises.¹¹²

¹⁰⁸ Similar terms were used by officials involved in the production of other Indexes of state fragility, to describe the value of an ordinal ranking: author interviews, Washington DC, July 20 and 22, 2010.

¹⁰⁹ See above, The emergence of state fragility.

¹¹⁰ The author has identified eight different Indexes of state fragility, state failure, or state weakness, including the USAID Index.

¹¹¹ Latour, *Reassembling the Social* (n. 16), 187–9.

¹¹² My thanks to Bruce Carruthers for bringing this similarity with heat maps to my attention.

Panoramas of this kind, then, effect distributed cognition, by tying together different scales of information and different genre systems (theories of politics, models of political organization, measures of economic growth, ratings of government policy, subjective perception surveys) in order to make “an otherwise amorphous composite...into a thing that holds together in the imagination of government officials and the general public.”¹¹³ Quantification of inputs eases the way to this knitting together, especially where scales are large, and obscures the nature of the genres being linked together. The (messy, unpersuasive) world-making that lies beneath this discrete numerical knowledge-object is obscured, and its seemingly solid and transparent qualities engender what Sally Merry calls “knowledge-effects”: “Numerical measures produce a world knowable without the detailed particulars of context and history.”¹¹⁴ In the analytical narrative that accompanies the 2010 USAID report, the figures derived from the legitimacy and effectiveness scores of certain states are used to diagnose, and predict, the nature of their fragility under environmental stresses caused by climate change. Referring to the World Bank’s list of 42 countries most vulnerable to five kinds of climate threat, the USAID Index report contends that

it will be possible to utilize the [USAID] Alert List rankings to make distinctions between these states. The advantage to this approach is that the fragility rankings and instability risk rankings are *tied explicitly to concepts that are clearly defined and distinct*. As a result, leveraging these lists to make assessments about which states are most vulnerable to threats posed will shed additional light on the specific problems likely to be encountered in these countries... The discussion below will distinguish between countries at risk for climate change in terms of *whether their vulnerability is rooted in problems of effectiveness, legitimacy, high levels of instability or conflict risk, or combinations of these factors*... [Of the 42 identified originally in the World Bank report], only 13 are classified in the ‘highest’ or ‘high’ category for fragility [by the USAID Index]... Accordingly, ... these 13 countries are arguably the most vulnerable *because the relationship between government and society in these countries is extremely fragile within both the realm of legitimacy and effectiveness*.¹¹⁵

We have seen that the concepts underpinning the USAID Index are neither clearly defined nor distinct, but this assertion of methodological authority cannot be challenged without considerable background research into the Index’s method of calculation. The plausible claim, on the face of the document, of conceptual and empirical rigor underwrites the knowledge-effect generated by the assignment of “legitimacy” and “effectiveness” scores—a broad diagnosis about the nature of the relationship between “government and society” in countries ranging from India to Mexico is read off the scores, and interpolated into a narrative that connects the

¹¹³ Espeland and Stevens, “A Sociology of Quantification” (n. 85), 412. See also Bowker and Leigh Starr, *Sorting Things Out* (n. 16), 136.

¹¹⁴ Sally Engle Merry, “Measuring the World: Indicators, Human Rights and Global Governance,” *Current Anthropology* 52 (2011): S84.

¹¹⁵ USAID, “2010 Alert List: Fragility and Risk for Instability: Worldwide Rankings,” USAID, 2010, 24–5 (my emphasis).

scores with an inability to successfully manage the risks arising from climate change.¹¹⁶ A bad score is associated with a judgment about poor performance in managing risks of instability and disorder, which is understood to include at least the possibility of a threat to neighboring states and even donor states.

I have already argued that, in the absence of better conceptual specification, better measures, and a theory of aggregation, we cannot know what the legitimacy and effectiveness scores generated by the USAID Index actually measure, if anything. But the way in which the scores are discussed in the above example does point us towards an understanding of how these numbers are *used* and thus the meanings that they can acquire. In an article on contemporary practices of security risk calculus, Amoores notes that “contemporary risk calculus does not seek a causal relationship between items of data, but works instead on and through the relation itself.”¹¹⁷ What Amoores terms the “ontology of association” associates items of data with one another, to generate amalgams of data, “inferring across the gaps to derive a lively and alert new form of data derivative—a flag, map or score that will go on to live and act in the world.”¹¹⁸ Amoores’s empirical example is that of risk profiling used by British border control agents, in which data mining and analytics are deployed to derive a risk score or flag for individuals entering or exiting the United Kingdom. But the insight seems to me to capture something important about the way in which the USAID Index acts, once we see clearly that the Index’s concepts and measures cannot live up to its methodological aspirations. *In practice*, the Index associates disparate measures and datasets with one another and bundles them into amalgams called legitimacy and effectiveness scores, which are then bundled further into a composite called a fragility score. The numbers thereby generated are then associated further with global judgments about the quality and nature of political order in the territory. The numbers *stand in* for a judgment about a complex social reality (e.g., the “relationship between government and society”) but also *tie in* with beliefs, common-sense notions, and normative claims about what characterizes good and bad political orders and outcomes. The normalizing tendency of the concept of fragility, discussed above, is reinforced but also rendered less transparent by the process of quantification; it is hidden behind claims of methodological authority and elaborate numerical artifacts.

¹¹⁶ “To the extent that managing the consequences of climate change will require effective delivery of core services to populations, successful implementation of major policy changes, or regulation or management of critical natural resource endowments, the states highlighted [in red] are likely to be in the worst position for handling these new challenges. These states are particularly vulnerable to the consequences of climate threats because they currently manifest low levels of capacity for effective performance of government functions... In countries in which the legitimacy component of the fragility score is high, it can be expected that those governments will have greater difficulty marshalling the necessary resources to gain public approval for sweeping policy changes necessary to address climate threats.” *Ibid.*, 25–6.

¹¹⁷ Louise Amoores, “Data Derivatives: On the Emergence of a Security Risk Calculus for Our Times,” *Theory, Culture and Society* 28 (2011): 27.

¹¹⁸ *Ibid.*

Conclusion: Calculability and hierarchies of sovereignty

A rank ordering is a hierarchy. In the case of the USAID Index, it is a hierarchy of fragility, which is composed of a hierarchy of state effectiveness and legitimacy. In the contemporary discourse of international law, the idea of “effectiveness” has come to feature prominently in arguments about the circumstances in which a state’s sovereignty may be suspended or rendered defeasible. In the context of *jus ad bellum*, there is increasing support for the notion that a state which is unable to control non-state armed groups within its territory (perhaps because it lacks full territorial control, or its security apparatus are ineffective) may well be subject to a proportionate and necessary use of force by another state which suffers an armed attack from the non-state actor.¹¹⁹ This position remains controversial, and cannot be considered *lex lata*. But for our purposes the interesting point is that one of the concepts invoked as a touchstone for “inability” in this argument is “effectiveness,” and it is used in conjunction with “legitimacy” to categorize states as lying somewhere between a spectrum of strength and weakness. The position of the state on this spectrum becomes a relevant consideration in determining whether it is unable or unwilling to control the armed group within its borders:

It . . . therefore makes sense to conceptualize state strength not in dichotomous but continuous terms. States around the world vary greatly with regard to their ability to provide public goods such as legitimate rule, welfare, and physical security. At one end of the continuum we have strong states which perform well across the board, while at the other extreme we find entirely collapsed states. These are neither governed legitimately, nor can they provide for the physical security of their citizens, nor for their welfare. In between these extremes we will find many shades of gray, however. It is for instance rather common that a state performs reasonably well in various dimensions of public goods provision, but is nonetheless incapable of controlling armed elements operating on its territory. . . . In the area of military intervention, a trend of curtailing the sovereign privileges of states that fail to prevent the use of their territory for internationally wrongful acts has emerged.¹²⁰

In a more strident vein, John Yoo argues that a failure to intervene where states cannot govern effectively produces severe negative externalities, which international law ought to play a role in containing. Thus, the categorical protection of sovereignty irrespective of a state’s effective sovereignty is counterproductive, and international law should “construct a different set of rules that would encourage intervention in failed states.”¹²¹ Interestingly, Yoo refers to several indices of state failure in order to argue that, while there may be some disagreement about precise rankings, there is broad agreement on which states qualify as failed or failing states.¹²²

¹¹⁹ See for a clear and careful overview, Theresa Reinold, “State Weakness, Irregular Warfare and the Right to Self-Defence Post-9/11,” *American Journal of International Law* 105 (2011): 244.

¹²⁰ *Ibid.*, 249.

¹²¹ John Yoo, “Fixing Failed States,” *California Law Review* 99 (2010): 95–150.

¹²² *Ibid.*, 101–4.

Hierarchies of sovereignty have a long history in international law;¹²³ indeed, they are arguably constitutive of the emergence of a state-centric international legal order. In the two decades since the end of the Cold War, the re-emergence of arguments against a categorical idea of sovereign equality, and some notion of degrees of sovereignty, has been widely observed.¹²⁴ One of the most important vectors of this trend has been debates over humanitarian intervention and its would-be successor concept, the responsibility to protect (R2P). The founding document of R2P, and its successive articulation in reports of the Secretary-General, seek to de-emphasize the role of military intervention in part by stressing the idea that states themselves bear responsibility for maintaining their sovereignty through the effective discharge of governing functions. Preventing atrocity is achieved through the maintenance of an effective and legitimate state:

The responsibility to protect, first and foremost, is a matter of State responsibility, because prevention begins at home and the protection of populations is a defining attribute of sovereignty and statehood in the twenty-first century. . . . Responsible sovereignty is based on the politics of inclusion, not exclusion. This entails the building of institutions, capacities and practices for the constructive management of the tensions so often associated with the uneven growth or rapidly changing circumstances that appear to benefit some groups more than others.¹²⁵

A programmatic approach to implementing R2P thus insists on forms of “early warning” which can monitor domestic political dynamics within states,¹²⁶ in order to facilitate “timely and decisive” responses (such as diplomacy, aid, or armed force).

Techniques which purport to render the degree of political order within a territory calculable and reducible to a single number, complement these shifts in law and policy, and reinforce them. More empirical analysis would be required to know whether and how the USAID Index or other indicators are being used as part of judgments about a state’s level of effectiveness and legitimacy, and how these may become part of classificatory judgments concerning state’s degree of sovereignty. But it would not be a leap to speculate that rank orderings of fragility or failure may become part of these judgments. It is evident that the concept of state which underlies both the concept of fragility and concepts such as R2P are similar, and so an elective affinity between measures of fragility and conclusions about a state’s discharge of its responsibilities of sovereignty, seems likely. Bowker and Leigh Starr note that classification systems (and an index is a classification system) “coalesce

¹²³ See Benedict Kingsbury, “Sovereignty and Inequality,” *European Journal of International Law* 9 (1998): 599–625; Edward Keene, *Beyond the Anarchical Society* (Cambridge: Cambridge University Press, 2002); Anthony Anghie, *Imperialism, Sovereignty and the Making of International Law* (Cambridge: Cambridge University Press, 2004).

¹²⁴ See Kingsbury, “Sovereignty and Inequality” (n. 123). See also Gerry Simpson, “Two Liberalisms,” *European Journal of International Law* 12 (2001): 537–71.

¹²⁵ United Nations Secretary-General, “Implementing the Responsibility to Protect,” January 12, 2009, UN Doc A/63/677, para. 14.

¹²⁶ *Ibid.*, Annex.

into working infrastructures when they are integrated into information systems,” and thus they provide “blocks against certain agendas and smooth [...] the road for others.”¹²⁷ It may be that quantifying degrees of fragility will become part of the process of calculating degrees of sovereignty, smoothing the road for the victory of a functionalist understanding of sovereignty.

¹²⁷ Bowker and Leigh Starr, *Sorting Things Out* (n. 16), 323–4.

PART II

INDICATORS, POWER, AND AUTHORITY IN GLOBAL GOVERNANCE

This page intentionally left blank

Re-Construction of Private Indicators for Public Purposes

*Katharina Pistor**

Introduction

In a world characterized by great uncertainties, there is a strong and legitimate need for reducing uncertainty by collecting, organizing, and classifying information in accordance with pre-established frameworks. The very first step—the collection of information—already requires analytical priors for determining which information to collect and which to ignore or discard. Organizing and classifying information are two additional steps to facilitate the interpretation of the collected information. The organization of information is often determined prior to actual data collection. The use of “blueprints” that set out what information is regarded as relevant and how to package it into symbols that can be easily understood and processed is the quintessential example of streamlining information to overcome “absorption uncertainty.”¹ Alternatively, information may be organized only after it has been collected. Once it has been collected and organized into clusters, and once these clusters have been labeled, conceptual boundaries are erected between different pieces of information, which may tend to exaggerate actual differences between indicators or indices. Such classifications become the new parameter, or blueprint, for collecting future information, thereby entrenching the choices made during the initial classification. In addition, they can and often do shape theories and concepts used for understanding the world. Absorption uncertainty is not only an issue for organization, but also for understanding and explaining complex social phenomena. As will be argued below, the construction of new rule of law and bureaucratic efficiency indices has been used to validate a new framing of concepts that were and continue to be much debated in academic circles.

* I would like to thank Bruce Carruthers, Nehal Bhuta, and other participants at the conference on “Indicators as a Technology of Global Governance” for helpful comments. All remaining errors are mine.

¹ James G. March and Herbert A. Simon, *Organizations* (New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons, 1958), 162.

The above implies that the identity and motivations of who collects information affect how information is processed and ultimately used. Knowing the source is therefore critical for a meaningful interpretation of information in general, and of information that has been aggregated and categorized in particular. This is hardly new. Yet, there are also strong interests to obscure the origins of information and classification schemes and to reify them as “facts” that speak for themselves and do not require further investigation into how they were produced.² Both producers and users of information and classification systems may have an interest in the reification of organizational choices into immutable “facts”: producers, because it lends economies of scale to the information systems they established; users, because it transforms partial views of the world into objective facts that can deflect critics and legitimate their use for normative purposes.

The main argument of this chapter is that the production of indicators as classified information and the process by which the origins of information and classification systems are obscured for normative purposes warrant particular attention in the analysis of indicators as governance technologies. The standardization of information for purposes of managing complex information is well known and has been widely practiced. The major difference in today’s proliferation of indicators as instruments of governance is their explicit use for governance purposes. When the original selection, organization, and classification of information is done for normative purposes, as in the World Justice Project’s (WJP) rule of law index,³ this purpose is sufficiently apparent to invite scrutiny and critique. However, when “hard data” from other sources are transposed into a normatively laden classification scheme, there is much less scrutiny about the origins of the data and the very process of transposition. After all, the bits and pieces that are used to construct the index come from an independent source and are therefore, supposedly, neutral.

The chapter develops this argument by tracing the history of the first generation of indicators for institutional quality, which have come to be known as “bureaucratic efficiency” or “rule of law” indices. It suggests that the private producers and users were quite aware of the origin and limitations of the information contained therein. Unlike some of the more recent private standard-setters discussed in Tim Büthe’s contribution to this volume, they were primarily concerned with information collection and organization for the purpose of risk management, less so with a normative agenda. Only with the discovery of these indicators by policymakers inside the World Bank and the publication of their application in respectable peer-reviewed journals was the process that led to their transposition into normative

² The contribution by Büthe in this volume illustrates this with respect to private standards and indicators. See Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011).

³ The Report that introduces the index states outright that its purpose is to advance the rule of law around the world. See World Justice Report, “The Rule of Law Index,” (Washington DC: World Justice Project, 2010). Note that the WJP was launched by the American Bar Association under then-President William H. Newcom, who now figures as the Founder, President, and CEO of the World Justice Project, <<http://www.worldjusticeproject.org/rule-of-law-index/>>. Newcom is the former General Counsel of Microsoft, and the WJP is financed in large part with money from Bill Gates, Sr.

benchmarks for policy purposes set in motion. The original producers did not protest against this abuse, as it lent credibility to their data and greatly expanded their use.

The origins and use of country risk indicators

Since the early 1970s several private firms have collected, organized, and categorized information about country risk profiles, which they sold for profit to multinational corporations to assess the riskiness of countries around the world.⁴ For example, Business Environment Risk Intelligence (BERI) was first compiled in 1972, and International Country Risk Guide (ICRG) in 1980.⁵ These were not the first attempts to quantify political and institutional factors for countries around the globe. The index of economic freedom compiled by Freedom House dates back to the 1950s, when the organization began to publish “balance sheets of freedom” for selected countries; it was systematized and expanded in 1972 under the leadership of Raymond Gastil. The technology for quantifying the unquantifiable was already established. The difference between Freedom House and BERI/ICRG was not only their different aspirations (defending “freedom” vs promoting business), but also their approaches to compiling data and making the compilation process transparent to their users. BERI and ICRG developed a substantially more fine-grained system and disclosed the different elements that went into the compilation of aggregate data. They were less interested in rendering a general verdict on a country—which was the rather explicit goal of Freedom House⁶—than in providing investors with information that they could parse for their own needs. Thus, the over 20 indicators that go into the compilation of the ICRG Political and Economic Risk indicators are based on a comprehensive model that allocates points to different components, thereby allocating different weights to them. The organization leaves it to its customers to select subcomponents to determine their own risk profiles. In other aspects ICRG follows more closely the path established by Freedom House than BERI. Like Freedom House, ICRG relies on in-house experts who compile and classify information they obtain from secondary sources—media, economic data available in public sources, and the like. By contrast, BERI uses surveys of their own correspondents in countries and regions and asks them to rank countries on a pre-determined scale. To ensure that correspondents follow similar standards, BERI

⁴ The business rationale was that companies would use information to determine the destination of their investments, but the data could be used for all kinds of purposes, and may have had only auxiliary relevance for corporate strategies.

⁵ The origins of the Business Intelligence (BI) indicators used by Mauro are not clear. See Paolo Mauro, “Corruption and Growth,” *The Quarterly Journal of Economics* CX, 3 (1995).

⁶ In light of this, it is rather surprising that the ideological underpinning of the Freedom House Index has not been discussed more widely. For a recent account along these lines see Diego Giannone, “Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case,” *Democratization* 17:1 (2010).

provides detailed guidelines on how to rank countries along the different dimensions in the questionnaire.

Each approach has its own methodological problems. Whereas in-house experts may develop a certain routine in compiling and classifying information, thereby ensuring greater commonality in the processing of this information, they are remote from the countries in question and have to rely exclusively on secondary information. Relying on a dispersed network of correspondents, in contrast, increases the variability in the interpretation of data, but ensures both a greater scope of and greater proximity to sources of information. Both approaches are ultimately subjective. They compile pieces of information into an indicator and ultimately an aggregate index.

Unfortunately, there is little information on how widely BERI or ICRG data were used by corporations and how much weight individual companies ultimately gave them when determining where, how, and how much to invest abroad. The fact that companies paid money for them, or that they survived over the years (although both were integrated into larger groups⁷), is not sufficient evidence that their services are valuable in the sense of being outcome-determinative.⁸ Indeed, they may play only an auxiliary function, or the role of a bargaining chip for obtaining tax holidays, government guarantees, or other sweeteners for choosing a particular investment destination. While several empirical studies suggest a correlation between some of the indicators discussed here and foreign direct investment (FDI) flows,⁹ this does not necessarily imply that BERI or ICRG were instrumental in these decisions.¹⁰

The Freedom House data had some impact on academic writings in political sciences, where they were frequently criticized¹¹ but nonetheless often used given the lack of alternative data.¹² In contrast, Business Intelligence (BI),¹³ BERI, and ICRG data were for years ignored by economists and policymakers alike, who hardly

⁷ BERI became part of the Economist Intelligence Unit, and ICRG is now published by the PRS Group.

⁸ See, however, Mauro, "Corruption and Growth" (n. 5), 684, who appears to justify the use of BERI data for measuring corruption with the fact that the database is costly.

⁹ John Hewko, "Foreign Direct Investment: Does the Rule of Law Matter?" *Carnegie Endowment for International Peace Working Papers*, No. 26 (2002).

¹⁰ Neither is it clear that the indicators adequately capture the levels of risk foreign investors might encounter, just that corporations actually relied on them. This leaves open the possibility that FDI flows were driven by indicators that suggest certain risk levels, which may or may not be accurate. The subsequent use of aggregate indices to test whether "institutional quality" determines investment decisions might then be measuring the responsiveness to ICRG or BERI, and not necessarily to actual conditions on the ground. Note also that some evidence points to the fact that foreign investors adapt to local standards, which sheds doubt on the notion that "good institutions" are a prerequisite for foreign investment flows. See Joel Hellmann, Geraint Jones, and Dani Kaufmann, "Far from Home: Do Foreign Investors Import Higher Standards of Governance in Transition Economies?" available at <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=386900> (2002).

¹¹ For a more recent critique and development of an alternative dataset, see T. Vanhanen, "A New Data Set for Measuring Democracy," *Journal of Peace Research* 37:2 (2000).

¹² The critique of the Freedom House Indicators has, however, given rise to a range of new datasets on political indicators. For an overview of datasets for comparative political research, see <<http://www.polisci.wvu.edu/polycy/psdata.html>>.

¹³ See above n. 6.

knew of their existence. Until the late 1980s, economists were preoccupied with macroeconomic growth theories and only slowly discovered the possible importance of politics or institutions for economic growth. To the extent they sought to account for political factors, they were satisfied with general measures of political stability. As late as 1991, Robert Barro used the number of coups per year and the number of political assassinations per capita as indicators of political stability.¹⁴ In reviewing his work, Stephen Knack and Philip Keefer¹⁵ criticize these measures as too crude to account for the different forms political instability can take. Indeed, dictators need not be overthrown or assassinated to create an atmosphere of instability for citizens and foreign investors alike.

In short, the country risk data that later formed the basis for assessing the institutional quality of countries around the globe were designed for quite different purposes: they were supposed to provide foreign investors with some background information about destination countries of their business. Moreover, they were around for years before development economists and policymakers in multilateral organizations discovered and found new applications for them. What needs to be explained therefore is not so much why they exist and how they were created in the first place, but their transposition to new applications, and in particular their transformation into measures of institutional quality and use in economics and for policy purposes.¹⁶

From subjective measures to normative claims about institutional quality

In 1995, two articles were published in important, peer-reviewed economics journals that used BI, BERI, and ICRG data to help explain economic growth. The timing coincides with the expansion of empirical economics since the late 1980s and a greater recognition of the importance of institutional variables for explaining economic growth.¹⁷ Only two years earlier, Douglas North was awarded the Nobel Memorial Prize in Economic Sciences for his work on institutions and

¹⁴ Robert Barro, "Economic Growth in a Cross Section of Countries," *The Quarterly Journal of Economics* 106:2 (1991).

¹⁵ Stephen Knack and Philip Keefer, "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," *Economics and Politics* 7:3 (1995).

¹⁶ For a comprehensive review of the increasing use of governance indices in the economics literature and problems associated with different indices, including those discussed in this chapter, see Andrew Williams and Abu Siddique, "The Use (and Abuse) of Governance Indicators in Economics: A Review," *Economics of Governance* 9:2 (2008).

¹⁷ Elevating "institutions" to the level of primary targets of policy intervention based on the assumption that institutional change would in fact make a difference itself amounted to a considerable shift in thinking. See Adam Przeworski and Fernando Limongi, "Political Regimes and Economic Growth," *Journal of Economic Perspectives* 7:3 (1993). See also Nehal Bhuta, "Democratization, State-Building and Politics as Technology," in H. Charlesworth, B. Bowden, and J. Farral (eds), *Great Expectations, the Role of International Law in Restructuring Societies* (Cambridge: Cambridge University Press, 2009).

institutional change in economic history.¹⁸ Both trends created a demand for more data. Applied empirical economics had largely relied on readily available macroeconomic data, whereas institutional research was either case-based or relied on formal models with some illustrations. If applied empirical economics using “large n-studies” was to make inroads into the new institutionalism, it needed institution-based data. The discovery of country risk data compiled by private consultancies provided them.

Paolo Mauro reports that he hand-compiled data from BI in the archives of the Economist Intelligence Unit (which had earlier acquired BI) in New York and used them to put to a test propositions about the likely impact of corruption on growth.¹⁹ Knack and Keefer published an article in the same year that uses both BERI and ICRG and pitches them against objective indicators of political stability as well as the indicators on political freedom compiled by the Freedom House.²⁰ Both papers were published in peer-reviewed economics journals—the *Quarterly Journal of Economics* (*QJE*), and *Economics and Politics*, respectively. All three authors had a keen interest in policy issues. Keefer was already at the World Bank. At the time, Knack worked at IRIS, a think tank on institutional development and economic growth at the University of Maryland, which was headed by Mancur Olson, who is credited as one of the founders of rational choice theory.²¹ Knack joined the World Bank four years later. Mauro joined the International Monetary Fund (IMF) the same year he published his *QJE* paper.

The research and policy interests of all three authors coincided with the new agenda of the Washington-based multilateral organizations—the Washington Consensus. It proclaimed that price stabilization, liberalization of trade and investment, and privatization—that is, policies that emphasized the primacy of markets—were deemed the engines of growth.²² The Washington Consensus was informed by a particular interpretation of previous policy experiences, especially in Latin America.²³ It could and did easily align itself with theories that assert the importance of markets and private property for long-term economic growth, such as those advanced by Douglas North.²⁴ However, beyond the historical case studies

¹⁸ <<http://www.nobelprize.org/nobel-prizes/economics/laureates/1993/>>.

¹⁹ See Paolo Mauro, “Corruption and Growth” (n. 5). Mauro asserts in this paper that to the best of his knowledge he was the first researcher to do this.

²⁰ See Knack and Keefer, “Institutions and Economic Performance” (n. 15).

²¹ See in particular Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press, rev. edn, 1971).

²² These policies have more recently come into doubt at the World Bank itself. See, for example, World Bank, *Economic Growth in the 1990s: Learning from a Decade of Reform* (Washington, DC: 2005).

²³ John Williamson, “What Washington Means by Policy Reform,” in John Williamson (ed.), *Latin American Adjustment: How Much Has Happened?* (Washington, DC: Institute for International Economics, 1990).

²⁴ Douglass Cecil North, *Structure and Change in Economic History* (1st edn, New York: Norton, 1981).

about economic growth and development in the United States, which North offered, there was little hard empirical evidence to support the notion that the asserted relation between private rights, markets, and growth held true around the world. The new applied institutional economics, which was formed by Mauro, Knack and Keefer, and others, provided this evidence. The data gap was filled by using indicators that had been compiled by private firms for multinational corporations to assess and manage their exposure to political risk in foreign markets. Critically, the indicators were not simply used in their original form; they were transposed from descriptive indicators about investment conditions into normative tools for policy reforms. This was done by clustering them into composites that received the economists' "efficiency" stamp of approval. Moreover, the re-labeling rather than the original construction of the indicators ensured consistency with the theories that were put to a test. The validity of the data and indicators was not judged by the soundness of the underlying theories, sampling techniques, or conceptual clarity, but in statistical terms by their explanatory power in support of the hypotheses that were tested. This gave them credibility as "objective" measures, notwithstanding the fact that they were built on subjective indicators that were originally collected for different purposes.

The papers by Mauro and Knack and Keefer offer detailed information about this transposition—making it all the more surprising how much of this history was subsequently lost in translation. Mauro prefaces his analysis with the opening statement that "[a]lthough most economists would probably agree that efficient government institutions foster economic growth, the magnitude of these effects has yet to be measured."²⁵ He sets out to fill this gap by selecting nine indicators from among the 56 risk factors compiled by BI, which he declares to be proxies for "institutional efficiency." This concept, however, is never explained. There is, of course, more than one concept of efficiency. More importantly, the concept is typically used to assess outcomes. Institutions may shape these outcomes, but how exactly institutions influence outcomes is context-specific. It is therefore not clear how a set of institutions can be deemed efficient without probing into the context in which they operate. Invoking this concept in this fashion is a critical step in framing the debate about institutions. It effectively transforms the descriptive data of BI into a value judgment of social systems.

Further, Mauro justifies the selection of only nine variables by arguing that they can be clearly separated from macroeconomic variables (which he seeks to explain and therefore cannot use as independent variables), and that they are general in nature, i.e., that they are of interest to any business, whether foreign or domestic. The latter criteria is not consistently applied, as can be seen from the Bureaucracy/Red Tape indicator in Table 7.1, which asks only about foreign firms' experience with red tape, not that of local investors or entrepreneurs. Mauro is also oblivious to the fact, or at least fails to problematize, that those surveyed were foreigners who might place greater weight on certain institutional arrangements

²⁵ Mauro, "Corruption and Growth" (n. 5), 682.

Table 7.1 Mauro's indicators of institutional efficiency

Indicators	Definition	Cumulative indices*
Political change— Institutional	Possibility that the institutional framework will be changed within the forecast period by elections or other means	PS
Political stability— Social	Conduct of political activity, both organized and individual, and the degree to which the orderly political process tends to disintegrate or become violent	PS
Probability of opposition group takeover	Likelihood that the opposition will come to power during the forecast period	PS
Stability of labor	Degree to which labor represents possible disruption for manufacturing and other business activity	PS
Relationship to neighboring countries	This includes political, economic, and commercial relations with neighbors that may affect companies doing business in the country	PS
Terrorism	The degree to which individuals and businesses are subject to acts of terrorism	PS
Legal system, judiciary	Efficiency and integrity of the legal environment as it affects business, particularly foreign firms	BE
Bureaucracy and red tape	The regulatory environment foreign firms must face when seeking approvals and permits; the degree to which it represents an obstacle to business	BE
Corruption	The degree to which business transactions involve corruption or questionable payments	BE

Source: Paolo Mauro, "Corruption and Growth," *The Quarterly Journal of Economics* CV, 3 (1995). 684.

*BE denotes "bureaucratic efficiency"; PS "political stability."

than domestic agents who may have access to alternative *modi operandorum* not easily accessible to foreigners.

Table 7.1 lists the nine indicators that were included in the study, their definition in the BI database, and their clustering into aggregate indices by Mauro.

Mauro then proceeds to analyze the correlation coefficient between the chosen indicators. He finds that they are all highly correlated with one another, and in statistically significant terms. To deflect criticism of multi-collinearity, he compiles them into cumulative indices. Perhaps a more appropriate remedy would have been to drop most of them (at the likely cost of reducing the explanatory power of the model), because the high correlation could also be viewed as questioning the distinct informational content of each indicator. By adding them up he increases the risk that unobserved variables—such as that those collecting information may

have been influenced by the general underperformance of the economy—are compounded.²⁶

The addition of highly selective indicators from the original database produces two compound indices: the Bureaucratic Efficiency Index (BE) and the Political Stability Index (PS). Mauro uses judiciary and red tape, as well as corruption, to construct BE, asserting that this index is a better measure of corruption conceptually than the corruption indicator BI measures directly. However, he does not actually perform a horse race between the two, or explain why the specific corruption indicator is thrown into the mix.²⁷ The compound index enables him to rank countries on a scale from 1–10.²⁸ He then correlates BE with the PS that comprises the remaining six variables. Once again, there is little explanation for why BI's own political stability indicator is not employed individually, or why, for example, terrorism and the possibility of disruption of manufacturing by labor are given equal weight in the aggregate index. He finds—not surprisingly given the earlier results—that both indices are highly correlated. He guards himself against the critique that this renders the explanatory power of these indices useless by arguing that there are enough outliers that the measure remains valid—i.e., countries that perform well on PS but low on BE: “Thus, even though bureaucratic efficiency and political stability are positively and significantly correlated, there is a wealth of information in the bureaucratic efficiency indices that can be used to analyse the determinants of investment and growth.”²⁹ Unfortunately, what exactly this wealth of information is and how one should unearth it after the indicators have been compounded remain unclear.³⁰

The final results of the regression analysis do not further concern us here—suffice to say that Mauro finds a negative correlation between BE and investment and between BE and growth (albeit a weaker one), and uses instrumental variable techniques to assert the direction of causality from institutions to economic outcomes. The lasting contribution of his paper has been the creation of a new index and one that would prove powerful for policy purposes: “If Bangladesh were to improve *the integrity and efficiency of its bureaucracy* to the level of that of Uruguay (corresponding to a one-standard-deviation increase in the bureaucratic efficiency index), its investment rate would rise by almost five percentage points, and its yearly GDP growth rate would rise by over half a percentage point.”³¹ The BI indicators were a critical tool for accomplishing this feat—but only after they were reorganized into cumulative indices and were given new meaning consistent with pre-ordained

²⁶ To be sure, Mauro discusses that subjective indicators are prone to these kinds of biases, but ultimately concludes that their benefits, namely their more detailed account of institutional quality, compensates for that. Moreover, he employs instrumental variables as a statistical strategy to correct for this. See *ibid.*, 690.

²⁷ *Ibid.*, 686.

²⁸ *Ibid.*, Table 7.1 at 687.

²⁹ *Ibid.*, 689.

³⁰ Note that similar issues have been raised with regard to the Freedom House Index. See John Hanson II, “Proxies in the New Political Economy: Caveat Emptor,” *Economic Inquiry* 41:4 (2003).

³¹ *Ibid.*, 705 (emphasis added).

theories: big government is bad for growth. Remarkably, in a single swoop, the empirical data that were reconstructed to fit the theory were also used to validate it.

The paper by Knack and Keefer follows a similar pattern. They also assert economic theory as given and the empirics turn into a hunt for data that are best able to confirm these theories. The very first sentence of the paper—once again citing North—establishes this agenda: “Few would dispute that the security of property and contractual rights and the efficiency with which governments manage the provision of public goods and the creation of government policies are significant determinants of the speed with which countries grow.”³² Note that they do not state that these are largely shared hypotheses, but assert the factual correctness of this statement. They too bemoan the lack of empirical data to prove this assertion and set out to examine how well different datasets explain it. As in the paper by Mauro, indicators are re-conceptualized and compiled into aggregate indices. Knack and Keefer use BERI as well as ICRG for this purpose, because they provide “detailed ratings for large samples on disaggregated dimensions of property rights that are closely related to those institutions emphasized by North (1990), Weingast (1993), Olson (1982) and others.”³³

It is of course legitimate to look for data that capture key elements of a given theory. But it is highly problematic to re-frame data to accomplish this task. As Knack and Keefer explain, ICRG’s “expropriation risk” and “rule of law”³⁴ are “interpreted here for the security of property rights and contract rights”; and “repudiation of contracts” “is another indicator for contract enforcement,” as well as an “indicator for government credibility.”³⁵ None of these new meanings is further defined; neither is there a hint that the transposition of descriptive data into normative judgments for policy purposes is even acknowledged. Similarly, “Corruption” and “Quality of the Bureaucracy” are used as proxies for the efficiency of government services. Tables 7.2 and 7.3 list the indicators Knack and Keefer selected for their purposes, the original definition of these indicators by ICRG and BERI, respectively (as summarized by Knack and Keefer), and their re-interpretation of these indicators.

One of the more remarkable transmutations of meaning is the re-labeling of ICRG’s “law and order index” as the “rule of law index.” ICRG’s law and order indicator already mingles “rule-based dispute resolution” with “orderly transfer of political power,” which arguably refer to very different social and political phenomena. It apparently was meant to assess whether disputes of any kind are resolved in a rule-based fashion or by use of violence. If anything, it should be restated as “rule *by* law.” Re-defining law-based behavior as rule *of* law adds an important normative aspect to the agenda. Rule of law is a complex and ambiguous concept, the meaning of which has been disputed for millennia.³⁶ As Waldron has argued, the very

³² North, *Structure and Change in Economic History* (n. 24), 207.

³³ Knack and Keefer, “Institutions and Economic Performance” (n. 15), 210.

³⁴ Note that ICRG denotes this as law and order.

³⁵ Knack and Keefer, “Institutions and Economic Performance” (n. 15), 211.

³⁶ J.M. Kelly, *A Short History of Western Legal Theory* (Oxford: Clarendon Press; New York: Oxford University Press, 1992).

Table 7.2 Knack and Keefer's interpretation of ICRG indicators

Indicator	ICRG definition	K&K interpretation
Quality of bureaucracy	High scores indicate "autonomy from political pressure" and "strength and expertise to govern without drastic changes in policy or interruptions in government services"; also existence of an "established mechanism for recruiting and training"	Efficiency of government services
Corruption in government	Lower scores indicate "high government officials are likely to demand special payments" and "illegal payments are generally expected throughout lower levels of government" in the form of "bribes connected with import and export licenses, exchange controls, tax assessment, policy protection, or loans"	Efficiency of government services
Rule of law	This variable "reflects the degree to which the citizens of a country are willing to accept the established institutions to make and implement laws and adjudicate disputes": higher scores indicate "sound political institutions, a strong court system, and provisions for an orderly succession of power"; lower scores indicate "a tradition of depending on physical force or illegal means to settle claims"; upon changes in government in countries scoring low on this measure, new leaders "may be less likely to accept the obligations of the previous regime". <i>Note: original variable name in ICRG is "law and order tradition"*</i>	Security of property and contract rights
Expropriation risk	Assessment of risk of "outright confiscation" or "forced nationalization"	Security of property rights
Repudiation of contracts by government	Indicates the "risk of a modification in a contract taking the form of a repudiation, postponement, or scaling down" due to "budget cutbacks, indigenization pressure, a change in government, or a change in government economic and social priorities"	Security of contract rights

Source: Stephen Knack and Philip Keefer, "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," *Economics and Politics* 7: 3 (1995), Appendix 1.

* Italics added by author.

essence of the rule of law is that it is a contested and contestable concept.³⁷ This, however, is not how Knack and Keefer use it; instead they invoke the aura of authority and respectability the concept—which subsequently has become a core ingredient of "good governance" as understood and propagated by the World

³⁷ Jeremy Waldron, "Is the Rule of Law Essentially a Contested Concept (in Florida)?" *Law and Philosophy* 21:2 (2002).

Table 7.3 Knack and Keefer's interpretation of BERI data

Indicator	BERI definition	K&K interpretation
Bureaucratic delays	Speed and efficiency of the civil service, including processing customs clearances, foreign exchange remittances, and similar applications	Corresponds to ICRG efficiency of government services
Nationalization potential	Measures risk of "expropriation for no compensation" and "preferential treatment for nationals"	Corresponds to ICRG security of property rights
Contract enforceability	Measures the "relative degree to which contractual agreements are honored" and "complications presented by language and mentality differences"	Corresponds to ICRG security of contract rights
Infrastructure quality	Assesses "facilities for and ease of communication between headquarters and the operation, and within the country," as well as quality of transportation	"[S]ome approximation" to the efficiency with which government services are provided

Source: Stephen Knack and Philip Keefer, "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," *Economics and Politics* 7: 3 (1995).

Bank's governance indicators³⁸—has gained in Western parlance. By codifying a measurable "rule of law index," they essentially deny the concept's contestability, and in Waldron's terms, thereby its core meaning.

The generalization of "contract repudiation by governments" into the much more general "security of contract rights" also stretches the meaning of an indicator that was specifically designed to assess the risk that contracts foreign investors enter into with host governments might be repudiated. It may well be the case that governments that repudiate contracts with foreign parties also care little about upholding contractual rights among private parties, but that is an empirical question, not one that should be decided at the level of indicator construction, or rather re-construction. The transformation of the BERI "contract enforceability" measure into a "security of contracts" indicator is even more problematic. The original indicator quite obviously tries to capture cultural variables that have little to do with the abuse of government powers or the lack of respect for private property rights.

The mutation of "infrastructure quality" into "efficiency of government services" in the BERI data is particularly disturbing. Low infrastructure quality may, but need not, be a function of the quality of government services. It may also be a function of poverty, i.e., of the simple lack of resources. If the concept of efficiency

³⁸ For a comprehensive discussion of these indicators, see Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi, *The Worldwide Governance Indicator's Project: Answering the Critics* (Washington, DC: The World Bank, 2007). A critique of the lack of a theoretical foundation for these measures can be found in Melissa Thomas, "What Do the Worldwide Governance Indicators Measure?," *European Journal of Development Research* 22:1 (2010, 31–54).

has any meaning at all, it refers to the efficient *allocation* of resources, not their availability or initial assignment, about which the efficiency paradigm has little to say.³⁹

In the next step Knack and Keefer conduct a horse race between ICRG and BERI on one hand, and the Freedom House index as well as the objective measures for political stability that Barro had used earlier on the other. They find that the “more specific” ICRG and BERI indicators (in the interpretation given to them by Knack and Keefer) “perform substantially better” than the Freedom House or the more fact-based political stability data Barro used. Specifically, in the growth model used for the regression analysis, the former’s explanatory power is greater than the latter’s. That, of course, does not mean that these are the better measures of institutions or institutional quality. It also turns on its head the purpose of testing theories empirically: the validity of the data or indicators used should not be asserted by their explanatory power in regression models, otherwise theories cannot ever be falsified.

From academic validation to public endorsement

The academic validation of the private risk indicators in peer-reviewed journals enhanced their usages in academic circles.⁴⁰ The two publications each render plus or minus 3,000 citations, respectively on *Google Scholar*. This speaks for a remarkable dissemination of the papers and the data used therein in academic circles broadly defined. The impact in policy circles has also been impressive. ICRG gets 169 hits on the IMF web page, and “corruption and growth”—the title of Mauro’s paper—gets 57 alone, which likely understates their actual influence on policy-makers. The data and concepts that were first developed in these papers have been deeply influential in the World Bank’s project on “good governance,” which led to a much more expansive aggregation of data compiled from various sources.⁴¹

In the hands of organizations with the mission to promote economic growth and development, the newly created aggregate indices were utilized for assessing the quality of institutions in countries around the globe and, critically, for deriving policy prescriptions. This was possible only because the origins of the data and their re-interpretation in the construction of the indices had been obscured. The original indicators were meant to inform foreign investors about the risk they might face when entering these markets. They were not concerned with the root causes of

³⁹ Lawrence H. White, “Can Economics Rank Slavery against Real Labor in Terms of Efficiency?,” *Politics, Philosophy and Economics* 7:3 (2008).

⁴⁰ The importance of authority derived from epistemic communities is also emphasized in Kevin Davis, Benedict Kingsbury, and Sally Engle Merry, “Indicators as a Technology of Global Governance,” *Law and Society Review* 36: 1 (2012): 71–104.

⁴¹ Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi, “Governance Matters IV: Governance Indicators for 1996–2004,” in *Policy Research Working Papers* (Washington, DC: World Bank, 2005); Daniel Kaufmann, Aart Kraay, and Pablo Zoido-Lobaton, “Aggregating Governance Indicators,” *Policy Research Papers* (1999).

these risks, much less with strategies to treat them. By converting these data into measures of institutional quality, the impression was created that improving the very same institutions these indicators purportedly measured would render the desired economic outcomes. Yet, as has been discussed above, none of the original data gave much information about specific institutions—notwithstanding Knack and Keefer's assertion that they were much closer proxies than objective indicators, or those of the Freedom House. Rather, the proximity of the institutional variables to the theories that were tested was made possible only by an affirmative act of re-interpretation.

Even if the data did capture information about real institutions, and even if the use of instrumental variables and other statistical techniques established a clear causal link between “bad institutions” and “good” economic outcomes, the implied possibility that reforming institutions judged as inferior would produce the desired outcome requires another leap of faith. It assumes that institutions are static structures,⁴² the characteristics of which are well understood. Somewhat contradictorily, it also assumes that institutions can be changed by fiat irrespective of underlying social, cultural, or political conditions. Finally, if it could be shown that the best performing countries relied on a specific set of institutions for their success, it does not follow that other countries can rely on the same set of institutions. Nor do these data say anything about what it would take to actually reform them. None of the empirical studies can truly claim otherwise. They use cross-sectional rather than time series data, and therefore have nothing to say about how individual countries improve their economic performance over time. Even those studies that attempt time series analyses have to concede that institutional variables for a large sample of countries are available only since 1980—and for most countries the data suggest that they changed only marginally over this time span.⁴³ Instead, the data were used to justify policies that had already been pronounced.

Even though much of this is well known, it has not diminished the power of these indices in policy circles. Their major advantage is that they support the ideological priors of the multilateral organizations, offer “hard data” to demonstrate their “truth,” and, in a conceptual framework that is built on simple causal relations between institutions and economic outcomes, can be easily translated into policy recommendations. The same organizations have also sponsored new research

⁴² This notion is supported by North's assertion that institutions are “path dependent” and slow to change. See Douglass Cecil North, *Institutions, Institutional Change, and Economic Performance* (Cambridge; New York: Cambridge University Press, 1990). More recent research on institutional change, however, has revealed that gradual change occurs frequently and can have far-reaching implications. See James Mahoney and Kathleen Thelen, “A Theory of Gradual Institutional Change,” in James Mahoney and Kathleen Thelen (eds), *Explaining Institutional Change* (Cambridge: Cambridge University Press, 2010).

⁴³ The construction of historical data on the quality of institutions this required is in itself a highly problematic undertaking that requires a much higher level of familiarity with specific institutions and the context in which they operate and change than these data aggregation efforts can possibly provide. For a critical assessment of historical databases, see only Evan S. Lieberman, “Bridging the Qualitative–Quantitative Divide: Best Practices in The Development of Historically Oriented Replication of Databases,” *Annual Review of Political Science* 13 (2010), 37–59.

projects to construct additional indices to serve similar purposes.⁴⁴ Indeed, the academic validation, if not creation, of datasets is a crucial step in legitimizing these indices by creating the appearance of objective science. This has also benefited the original producers of the datasets—the private risk consultancies that created them. ICRG includes links on its own website to academic publications and policy papers by the IMF,⁴⁵ offers discounts to academics, and lists databases created at academic institutions that refer to or are built on their indicators.⁴⁶ This suggests that ICRG has viewed as beneficial the transposition of its own datasets into different domains. The public endorsement of private data, even in their altered form, has enhanced their credibility.

Lessons for indicators as governance technology

The lessons one might draw from the above analysis are simple: the collection, organization, and classification of information into indicators as such is a common and useful exercise. It reduces the complexity of information and facilitates its interpretation. The use of similar information for policy purposes or as governance tools is also both possible and desirable. As long as the process of information collection and processing is transparent and contestable, individual indicators or aggregate indices can be used for appraising the need for, as well as the success of, policy interventions. The problems with the construction of the institutional quality variables are rooted in the ideology that motivated their selection, reconstruction, and use without much transparency.⁴⁷

What, then, would a more meaningful use of indicators as a means to improve governance look like? One could start simply by going back to where BERI and ICRG began: offer data to end-users, which can be used in whatever fashion users see fit. For this purpose, raw data should be accessible, including information about sampling techniques. In addition, the purpose for which the data are used should be clarified. Just as multinational firms purportedly use ICRG, BERI, or similar data to help them determine the nature of the risks they are willing or able to accept in foreign markets, so should domestic and international policymakers be able to choose policy objectives and determine on a case-by-case basis what institutional reforms might be desirable to accomplish them. Finally, alternative datasets should be created and employed to challenge existing priors rather than to validate them. This would allow policymakers and affected constituencies to contest the analyses and policy conclusions presented under the guise of hard science.

⁴⁴ Most of the indices incorporated in the *Doing Business* database of the World Bank were constructed by academic economists at Harvard and elsewhere.

⁴⁵ The PRS Group—About Us, <http://www.prsgroup.com/AboutUs_Overview.aspx>.

⁴⁶ See <<http://www.prsgroup.com/AcademicTitles.aspx>> (last visited August 26, 2010).

⁴⁷ In a similar vein, the Freedom House Indicators' ideological bent has been re-enforced by their use to motivate funding of development projects by the Millennium Challenge Foundation and USAID. See Giannone, "Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case" (n. 6), 75.

Legal Yardsticks: International Financial Institutions as Diagnosticians and Designers of the Laws of Nations

*Terence C. Halliday**

Introduction

In the past 20 years, the legal departments of the world's most powerful international financial institutions (IFIs) have undergone a fundamental shift in emphasis. Once dedicated to servicing transactions conducted by the financial arms of their multilateral banks, IFI legal departments have emerged as powerful institution-builders of legal and regulatory systems, indeed, as "architects of states."¹ An integral element of this enterprise has been the calibration of legal systems—the development and application of legal yardsticks across the nations of the world. Through a combination of diagnostic instruments and programmatic extrapolations, IFIs have erected a technology of appraisal and application that embraces, in principle, every country, rich or poor and in between.

This movement towards a reliance on indicators by IFIs corresponds with a broader use of indicators as a technology of global governance.² Widely deployed by public and private entities, indicators purportedly simplify global and national governance, reduce the burdens of policymaking in complex national and global regimes, and lend authority to decision-makers everywhere and those who seek to influence them. But indicators themselves are permeated by the exercise of power—in their creation, their propagation, and their impact. If social indicators

* I appreciate the manifold contributions of my collaborators on our wider global norm-making projects, Susan Block-Lieb and Bruce Carruthers, and thank the participants at the NYU Indicators conference and the editors of this volume for their feedback. I gratefully acknowledge the research funding of the National Science Foundation (Grant SES-0214301) and the American Bar Foundation for aspects of this study.

¹ Terence C. Halliday, "Architects of the State: International Organizations and the Reconstruction of States in the Global South," *Law and Social Inquiry* (forthcoming), 2012.

² See Kevin Davis, Benedict Kingsbury, and Sally Engle Merry, "Indicators as a Technology of Global Governance," *Law and Society Review* 36:1 (2012): 71–104.

are the creatures of social scientists, they must also be subject to the same sorts of scrutiny they purport to exemplify.

This chapter analyzes the intricacies and traces of power within indicators developed by IFIs to diagnose and design national legal systems. It will advance five theoretical propositions. First, the use and form of indicators by any one IFI reflects the structure and dynamics of the ecology of international organizations in which it is embedded. Secondly, the use of indicators is driven by organizational imperatives that require IFIs to take on enormous tasks for which they are under-resourced. Thirdly, an implicit and sometimes explicit epistemological tension among professions within IFIs impels their legal departments to diagnose national legal systems with technologies drawn more from the social sciences and finance than law. Fourth, because the competitiveness of IFIs as global norm-makers depends upon their legitimacy, they are pressed to adopt representations of complex phenomena—legal systems—in forms that are acceptable to their core constituencies. Fifth, as a lever of legal change indicators are embedded in a dynamic temporal process of the recursivity of law which opens empirical and theoretical analysis of indicators and their fates so it can be informed by other episodes of legal change in a global context.

I develop these arguments on the basis of empirical research in two domains: a book on global norm-making by international trade law organizations (ITLOs) and a project of the Center on Law and Globalization on measuring law by IFI legal departments.³ These develop earlier research on the role of international organizations in the construction of international trade law, most notably for corporate bankruptcy and restructuring law.⁴ I shall rely principally in this chapter on the activities of four IFIs in the ongoing elaboration of norms for corporate bankruptcy regimes and commercial law more generally—the IMF, World Bank, European Bank for Reconstruction and Development (EBRD), and Asian Development Bank (ADB). We have previously reported on some aspects of the development of bankruptcy indicators in the late 1990s and early 2000s.⁵ Here I dig a little deeper, turn over the objects of analysis to perusal from another angle, and bring into the present the pursuit of measurement techniques for the purpose of systemic financial stability.

The chapter divides into two sections: the first describes in some detail the technologies themselves, their forms and functions; the second reflects theoretically on the significance of the empirical materials for understandings of global governance.

³ The book on ITLOs is jointly authored with Susan Block-Lieb, Fordham University, and focuses on the United Nations Commission on International Trade Law (maritime law, secured transactions law, insolvency law), UNIDROIT, and the Hague Conference on Private International Law, which variously cooperate and compete with each other over the development of norms for international and domestic trade. The project on IFI legal departments examines the variety of methods used by legal departments for evaluating substantive law and legal institutions for their respective members.

⁴ See particularly Terence C. Halliday and Bruce G. Carruthers, *Bankrupt: Global Lawmaking and Systemic Financial Crisis* (Palo Alto: Stanford University Press, 2009).

⁵ Terence C. Halliday, Susan Block-Lieb, and Bruce G. Carruthers, "Rhetorical Legitimation: Global Scripts as Strategic Devices of International Organizations," *European Socio-Economic Review* (2009): 1–36; Terence C. Halliday and Bruce G. Carruthers, *Bankrupt: Global Lawmaking and Systemic Financial Crisis* (n. 4).

Varieties of legal yardsticks

In the wake of national debt crises in the mid-1990s and the Asian Financial Crisis beginning in 1997, the world's leading clubs of nations, the G-7 and G-22, charged IFIs with the task of creating a global financial architecture that would forestall either national or regional financial collapse. Part of the packet of reforms that all developing and transitional countries were to consider included the construction of a national bankruptcy system for failing companies. The optimal system would have six elements: substantive and procedural corporate bankruptcy law; efficient and effective courts; a state agency to administer aspects of bankruptcies; an out-of-state administrative apparatus that would resolve debtor-creditor conflicts before they overwhelmed the courts; and profession(s) to provide the expertise to make all the previous elements function effectively. The logic behind IFI interventions was rather straightforward: if failing companies could be saved before they collapsed altogether, or if they could be liquidated in an orderly manner when debt burdens became unmanageable, then banks would not build up dangerous levels of non-performing loans, central banks would not become vulnerable to massive failure of domestic financial institutions, and national currencies and reserves would not place an entire nation's financial system in jeopardy. Robust national institutions would short-circuit potential regional crises that might even threaten the international financial system.⁶

So, in an ex post facto construction of crisis causation theory, if the so-called Asian Tigers—Indonesia, Malaysia, Thailand, South Korea, Taiwan—had built strong corporate bankruptcy systems inside their spectacularly successful trajectories of economic development, then many of the pressures on each nation's financial systems would have been observed long before the banking system began to topple. Good law, in short, would have forestalled national economic crisis and protected regional and global financial stability.⁷

The sprouting of corporate insolvency systems grew not only out of crisis in poorly institutionalized markets, but from the shifts from command to market economies that followed the collapse of the Soviet Union. All over Central and Eastern Europe the designers of market institutions insisted upon the building of mechanisms to handle businesses that did not survive market competition. The massive investment in the legal foundations of market institutions in the former USSR entailed the construction of corporate bankruptcy systems.

The two great regional convulsions of the 1990s therefore precipitated a scramble by IOs and IFIs to create legal norms, to measure deviation from observance of those norms, and to erect compensatory structures that would exemplify adherence to regional and global normative standards. Transitional and developing societies

⁶ G-22, "Report of the Working Group on International Financial Crises" (World Bank, 1998).

⁷ Gerald E. Meyerman, "The London Approach and Corporate Debt Restructuring in East Asia," in *Conference on Emerging Markets in the New Financial System: Managing Financial and Corporate Stress* (Upper Saddle River, NJ: Prentice Hall, 2000).

were targeted for thoroughgoing reviews and evaluations, not only in corporate bankruptcy law, but across some 12 components of a fully defined commercial and financial system.⁸ All this required yardsticks, so it was thought. Hence institution-building in domestic markets was accompanied, indeed stimulated, by transnational development of norms, diagnostic instruments, and implementation vehicles.⁹

European Bank for Reconstruction and Development

The EBRD was charged in 1990 with multilateral leadership of the vast efforts to reconstitute formerly command economies as fully functioning market economies. The EBRD differentiated itself from existing regional and global IFIs in five respects.¹⁰ First, the founders adopted a European consensus that standards of *political conditionality* mandated that assistance would only go to countries that made “clear commitments regarding the rule of law, respect for human rights, the establishment of multiparty systems, the holding of free and fair elections in the course of 1990, and economic liberalization with a view to introducing market economies.”¹¹ Secondly, the EBRD determined to focus on private-sector development and especially on small and medium-sized businesses (SMEs). Extant development theory and political ideology held that growth of the SME sector would provide an engine for economic growth and a vital condition for political democracy. Thirdly, the EBRD emphasized environmental issues. Fourth, the EBRD acquired “a flexible set of tools unprecedented in a single multilateral development bank.” And finally, the Bank would take on a significant role in advising governments about privatization and developing their private financial sectors. All these distinguishing goals turned on an agreement that 60 percent of EBRD funds would be directed to private financing and a maximum of 40 percent would be committed to public-sector projects.

The EBRD quickly became the single largest investor in Central and East European countries by leveraging its investment through partnerships with private commercial banks, other IFIs, and other multilateral institutions. Its private investment character ensured that EBRD staff became acutely aware of how credit markets operated in practice, since they needed to obtain security and at times realize their collateral when investments were at risk. Since a bank committed

⁸ For instance, the IMF and World Bank have developed some 12 diagnostic instruments related to financial stability. These Reports on the Observance of Standards and Codes (ROSCs) cover a range of national institutions thought to be integral to stable national financial systems: Data Dissemination; Fiscal Transparency; Transparency in Monetary and Financial Policies; Banking Supervision; Securities Regulation; Insurance Supervision; Payments and Settlements; Corporate Governance Accounting & Auditing; Insolvency and Creditor Rights; Anti-Money Laundering; and Combating the Financing of Terrorism.

⁹ See a complementary interpretation of IFI and IO global norm-making and evaluations in Halliday and Carruthers, *Bankrupt* (n. 4), ch. 3.

¹⁰ See Stephen Weber, “Origins of the European Bank for Reconstruction and Development,” *International Organization* 48: 1–38, esp. at 19–20.

¹¹ Weber “Origins of the European Bank for Reconstruction and Development” (n. 10), p.11, n. 29; [EC Commission document].

predominantly to private lending must depend upon collateral for loans and investments, the Legal Department undertook its first major lawmaking project on the law of secured transactions.¹²

This initiative coincided with the appointment of a new chief economist at the EBRD who pressed the Bank to examine closely the impact of EBRD financing and projects on the global economy of a country. The same logic applied to the Legal Department—to examine what impact law reforms had on the economy, and vice versa, and thus become more purposive about the institutional and behavioral dimensions of reforms. Not coincidentally, the more persuasively the Department could show the likely impact of law reform projects on a country's market, the better positioned the lawyers would be when presenting projects for approval before the EBRD Board.¹³

The program developed along four tracks.¹⁴ First, it sought to assess and evaluate laws, to create an inventory and database of laws, and to evaluate them against some standard. The Legal Transition Program launched Legal Indicator Surveys which systemically appraised substantive law and practice in all 26 EBRD countries of operation in core areas of commercial law, including insolvency.¹⁵ Secondly, the Programme undertook many project-specific, country-specific initiatives across the region. Thirdly, the EBRD joined with other IFIs, most notably the IMF, World Bank, and United Nations Commission on International Trade Law (UNCITRAL), to develop international standard-setting products. Although the EBRD itself did not propagate its own normative standard, it drew on its experiences in EBRD countries of operation to react to standards developed by other international institutions. And, finally, the Legal Transition Programme expanded its dissemination efforts by turning its Law and Transition newsletter into a fully-fledged journal, and by publicizing in print and at its website the evaluative surveys the Legal Department undertook in EBRD countries of operation.

Legal indicator survey

From 1997 to 2002 the Legal Department's primary instrument for evaluating legal systems was a legal indicator survey. It was premised on a basic socio-legal insight. On the one hand, a country's law could be appraised for its comprehensiveness or *extensiveness*, i.e., how well did it cover the principal topics that a "good" corporate insolvency law should embrace? On the other hand, extensive law on the books meant nothing without appropriate implementation. Thus the survey also measured *effectiveness*, i.e., whether "legal rules are clear and accessible and adequately implemented administratively and judicially."¹⁶

¹² Interview 2022. Senior officials in the Legal Dept, EBRD.

¹³ Ibid.

¹⁴ David S. Bernstein, "Process Drives Success: Key Lessons from a Decade of Legal Reform," *Law In Transition* (2002): 2–13.

¹⁵ Interview 2022; Bernstein, "Process Drives Success" (n. 14).

¹⁶ European Bank for Reconstruction and Development (EBRD), "Transition Report 1999: Ten Years of Transition," (London: EBRD, 2009), Annex 2.2, p. 1.

Countries were ranked on an ordinal scale for the extensiveness and effectiveness of their law: it might be comprehensive, adequate, barely adequate, inadequate, or detrimental. A country's commercial law would be rated "comprehensive" if "comprehensive legislation exists in all three areas of commercial law . . . [namely,] pledge, bankruptcy and company law. Legal rules closely approximate those of more developed countries" . . . and the law, institutions, and professions score highly on other attributes (Table 8.1). A country's commercial law would be given high rankings on "effectiveness" if:

Commercial laws are clear and readily ascertainable. Commercial law is well-supported administratively and judicially, particularly regarding the efficient functioning of courts, liquidation proceedings, the registration of shares and the orderly and timely registration of security interests (Table 8.1).

Countries were not only evaluated by their adherence to EBRD notional standards, but they were also rated numerically so their appraisals could be tracked over time and their relative standing vis-à-vis other countries could be publicly displayed. Rather than rely on the sole judgment of an academic lawyer or a leading practitioner or consultant, the EBRD conducted a survey of *perceptions* by practicing lawyers, academics, and other experts in a country. It then aggregated and scaled responses¹⁷ on a four-point scale, sometimes with pluses and minuses, to yield eight potential values. These were reported separately for extensiveness and effectiveness, but could also be summed or averaged for a general country score (Table 8.2). Scorecards or performance tables were published in the EBRD's publication, *Law in Transition*, which was widely circulated in justice and finance ministries across Central and Eastern Europe.

Commentary on the tables points to changes in performance over time. For example: "Kyrgyzstan's score has improved"; the "Czech Republic received a slightly lower rating"; "Bulgaria has had problems."¹⁸ Often the commentary lauds particular countries for their positive steps: "1994 Bulgaria Law provides this clear priority for secured creditors"; "Estonia and Croatia have established standards for trustees along lines advocated by the Bank." Negative examples may be adduced simply to show that standards have not been attained, as in the assessment that minority shareholders in Russia complain that they are not heard or cannot vote on reorganization plans. A stronger version of the negative example, however, obtains when it can be shown to produce adverse effects, such as the harm produced where Hungary chose an automatic trigger and the courts were flooded with filings.

¹⁷ Among the methodological weaknesses of the Legal Indicators Survey were (a) a failure to disclose numbers of respondents for any country; and (b) what scaling protocols were used; or, most egregiously, (c) the apparent arbitrary adjustment of scores so the EBRD itself made a judgment about "a given country to arbitrate among competing views of survey respondents" (Anita Ramasastry, "What Local Lawyers Think: A Retrospective on the EBRD's Legal Indicator Surveys," *Law in Transition* (2002), 14–30, 18.

¹⁸ The EBRD, Transition Report, 1999 (n. 16), Annex 8.1.

Table 8.1 Examples of maximum and minimum criteria for rankings of commercial law systems in Central and Eastern Europe for extensiveness and effectiveness**Classification system for legal transition indicators: commercial law****Extensiveness**

Minimum score = 1

Legal rules concerning pledge, bankruptcy, and company law are very limited in scope. Laws impose substantial constraints on the creations, registration, and enforcement of security over movable assets, and may impose significant notarization fees on pledges. Company laws do not ensure adequate corporate governance or protect shareholders' rights. Bankruptcy laws do not provide for certainty or clarity with respect to the definition of an insolvent debtor, the scope of reorganization proceedings, or the priority of distribution to creditors following liquidation. Laws in these substantive areas often have not been amended to approximate those of more developed countries and the laws that have been amended contain ambiguities or inconsistencies.

Maximum score = 4+

Comprehensive legislation exists in all three areas of commercial law that were the subject of this survey—pledge, bankruptcy, and company law. Legal rules closely approach those of more developed countries. These legal systems have a uniform (that is, centralized registration) system for taking and enforcing a security interest in movable assets and also provide for adequate corporate governance and protect shareholders' rights. In particular, the rights of minority shareholders are protected in the event of the acquisition by third parties of less than all of the shares of a widely held company. Bankruptcy law provides in a comprehensive manner for both reorganization and liquidation. Liquidators possess a wide variety of powers and duties to deal with the property and affairs of a bankrupt, including wide powers of investigation of pre-bankruptcy transactions carried out by the debtor. There are specialized courts that handle bankruptcy proceedings. Liquidators must possess certain minimum qualifications.

Effectiveness

Minimum score = 1

Commercial legal rules are usually very unclear and sometimes contradictory. The administration and judicial support for the law is rudimentary. The cost of transactions, such as creating a pledge over a movable asset, is prohibitive so as to render a potentially extensive law ineffective. There are no meaningful procedures in place in order to make commercial laws fully operational and enforceable. There are significant disincentives for creditors to seek the commencement of bankruptcy proceedings in respect of insolvent debtors.

Maximum score = 4+

Commercial laws are clear and readily ascertainable. Commercial law is well supported administratively and judicially, particularly regarding the efficient functioning of courts, liquidation proceedings, the registration of shares, and the orderly and timely registration of security interests.

Source: European Bank for Reconstruction and Development (EBRD), "Transition Report 1999: Ten years of Transition." (London: EBRD, 1999), p. 45.

Legal sector assessments

In 2002 the EBRD Legal Department changed course. Rather than rely on perceptions of multiple actors in a country, the EBRD commissioned consultants in various areas of substantive law to undertake a country-by-country analysis of law-on-the-books in 27 states where EBRD operated. For bankruptcy law, two

Table 8.2 Sample of cross-national rankings of CEE countries on commercial law indicators, 1998, 1999

Country	Legal transition indicators: commercial law					
	1999			1998		
	Overall	Extensiveness	Effectiveness	Overall	Extensiveness	Effectiveness
Albania	2	2	2-	2	2	2
Armenia	3-	4-	2	3	4	3
Azerbaijan	3-	3+	2	2	3	2
Belarus	2	2	2	2	2	2
Bosnia & Herzegovina	2-	2	1	1	2	1

Source: European Bank for Reconstruction and Development (EBRD), "Transition Report 1999: Ten years of Transition." (London: EBRD, 1999), p. 45.

eminent practitioners were retained by the Legal Department.¹⁹ The reports offer a systematic review of particular countries and a cross-sectional analysis of all countries. Both are facilitated by the ranking of countries, topic by topic, on a five-point scale of compliance with EBRD norms (from "very high" to "very low"). Thus information could be presented in several configurations: as numerical ratings on any given topic; as qualitative rankings by country; and as aggregate rankings by area of law or by country.

In the bankruptcy area, the first public report released in mid-2003 warrants that the broad contours of the five "core essentials" of an insolvency law²⁰ represent an international consensus.²¹ These five areas are in turn broken down into 23 major subject areas. A compliance score is created for each subject area, aggregated to a score for each of the five core areas, and aggregated yet again to provide an overall compliance score for each country. Table 8.3 shows variations across countries on a single criterion—that of "assets of a bankruptcy estate."

It should be noted that this ranking does more than implicitly derogate countries that fail to comply. The EBRD uses the poor results to pedagogic effect, because it employs the accompanying commentary to describe under each of the five main

¹⁹ Ron Harmer (an Australian lawyer and key drafter of Australia's corporate bankruptcy law in the 1990s, consultant to the Asian Development Bank, World Bank, European Bank of Reconstruction and Development; non-state observer delegate to UNCITRAL Working Group on corporate bankruptcy); Neil Cooper (English accountant and corporate restructuring specialist, former President, International Federation of Insolvency Practitioners, consultant to EBRD on bankruptcy systems in Central and Eastern Europe, non-state observer delegate to UNCITRAL Working Group on corporate bankruptcy).

²⁰ These covered five areas: (1) Commencement and Effects of Proceedings; (2) Assets of the Estate; (3) Creditors: Involvement and Treatment; (4) Reorganization Process; and (5) Terminal (bankruptcy/liquidation) Process.

²¹ These included the International Monetary Fund ("Orderly and Effective Insolvency Procedures" May, 1999), the World Bank ("Principles and Guidelines for Effective Insolvency and Creditor Rights Systems," April 2001), the Asian Development Bank ("Good Practice Standards for an Insolvency Regime" in Law and Policy Development at the Asian Development Bank, April 2000) and the UNCITRAL Working Group on "Legislative Guidelines for Insolvency Law" (at the time).

Table 8.3 EBRD assessment of countries' compliance with insolvency standards^a

B. ASSETS OF THE ESTATE

In this area the range of compliance scores is between "very low compliance" and "high compliance." The average compliance falls within the "low compliance" category. In this core area, there was **least compliance**. In this category, the countries achieving the different levels of compliance, in alphabetical order, are as follows:

Countries rated "very high"

NIL

Countries rated "high"

Croatia, Estonia, Serbia and Montenegro, Slovak Republic

Countries rated "medium"

Albania, Belarus, Bulgaria, Czech Republic, FYR Macedonia, Latvia, Moldova, Poland, Romania

Countries rated "low"

Armenia

Countries rated "very low"

Azerbaijan, Bosnia and Herzegovina, Georgia, Hungary, Kazakhstan, Kyrgyz Republic, Lithuania, Russian Federation, Slovenia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan.

^a Source: *Insolvency Sector Assessment, 2003*.

subject areas the principal shortcomings it has observed in the region, why they are problematic, and thus why they should be remedied. For example, regarding the assets of the estate—the second of the five areas—the EBRD notes that three problems recur in many of the countries: safeguarding of assets; provision of information; and avoidance of pre-bankruptcy transactions. Unlike the Asian Development Bank (ADB) report, however,²² the EBRD Assessment Survey rarely upbraids particular countries. Essentially it allows variations in results for the same country over time, or variations across countries on the same criterion, to shock, embarrass, shame, or prod countries into domestic reform programs that might, in practice, be supported by technical assistance from the EBRD or other IFIs, such as the World Bank.

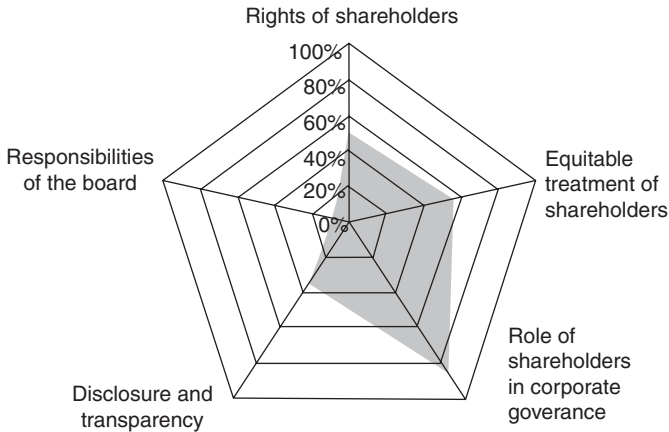
The Legal Department then converted its qualitative rankings into "spider graphs" (Figure 8.1), which report data country-by-country and often topic-by-topic. These dramatic visual representations of compliance concomitantly conveyed an appearance of scientific verisimilitude (see rankings from 0 to 100 on each criterion) and spatial indications of distorted or stunted legal systems. This "aesthetics" of indicators demonstrates a clear understanding of the value of presentation over and above its "science."²³

²² Ronald Harmer, an Australian lawyer who was a primary drafter of Australia's 1990s' reforms, was a co-author of the ADB Report and a consultant in the 2000s to both the ABD and EBRD's bankruptcy reforms.

²³ See also Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 2).

CHART 4

Quality of corporate governance
legislation in Ukraine



Note: The extremity of each axis represents an ideal score in line with international standards, such as the OECD Principles of Corporate Governance. The fuller the “web,” the more closely corporate governance laws of the country approximate these standards

Source: EBRD Corporate Governance Sector Assessment, 2004

Fig. 8.1 “Spider graph” of corporate governance compliance in the Ukraine

Legal indicator survey based on case studies

In 2004, the EBRD changed its methods of appraising effectiveness of legal change. It charged its consultants to present local lawyers—usually one firm per country—with vignettes of typical commercial transactions and to ask them to report on how well they actually worked in practice—how long each step took, whether courts acted as they were intended to do, whether execution of court orders actually took place, and the like. Again, criteria were quantified and represented in the graphs, which combined comparisons across major issues within an area of substantive law and cross-country comparisons, as the 2005 report on “concessions law-in-practice” reveals (see Figure 8.2). This mode of numerical and graphic representation continues to the present.

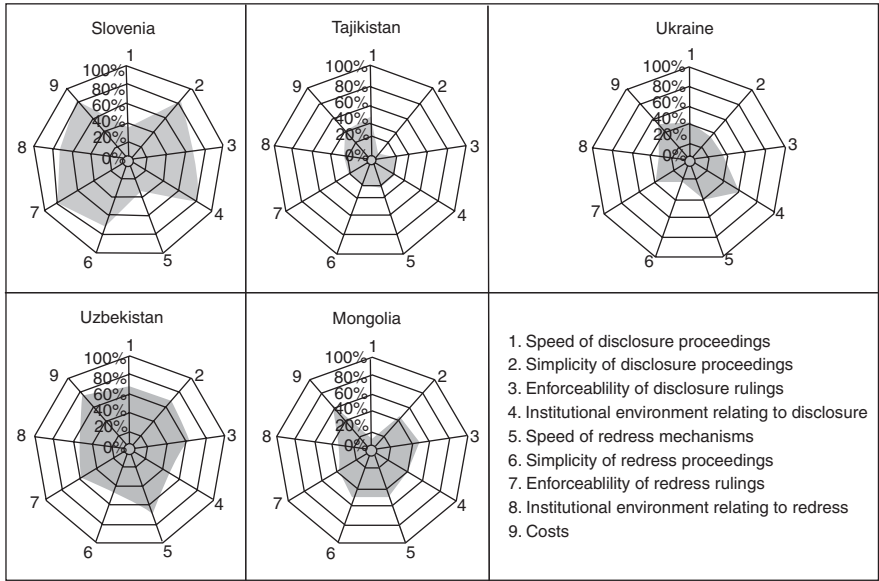


Fig. 8.2 Selected EBRD ratings of effectiveness of corporate governance legislation in case study scenarios

Source: European Bank for Reconstruction and Development (EBRD), “Transition Report 1999: Ten years of Transition,” (London: EBRD, 1999).

Asian Development Bank

The Legal Department of the ADB entered the enterprise of systematically evaluating national laws following the Asian Financial Crisis—some eight years after the EBRD—although it had a long-standing law and development program.²⁴ In the wake of the crisis, it commissioned an Australian law firm to codify good practice standards for an effective bankruptcy system and then to use those as a yardstick to appraise systematically 11 Asian countries, some developed (Hong Kong, Japan, Singapore, Taiwan, Korea) and some developing (India, Indonesia, Malaysia, Pakistan, Philippines, Thailand).

The ADB consultants (incidentally led by Ron Harmer, who later undertook national country assessments for the EBRD) crafted 33 good practice standards which they characterized as indicative of “advanced” or “modern” or “effective” insolvency systems.²⁵ On each standard, a country was ranked according to the following questions:

²⁴ Katharina Pistor and Philip A. Wellons, *The Role of Law and Legal Institutions in Asian Economic Development: 1960–1995* (Oxford University Press, 1999).

²⁵ Asian Development Bank, *Law and Development at the Asian Development Bank* (Asian Development Bank, 1999); Asian Development Bank, “Report on Insolvency Law Reforms in the Asian and Pacific Region,” *Law and Policy Reform Bulletin* (2000): 1: 10–86.

Table 8.4 Sample rating table on insolvency good practice standards

Good Practice Standards 8.1 and 8.2

(1) The law should prescribe, as fully as possible, for the provision of relevant information concerning the debtor.

(2) It should also provide for independent comment and analysis of the information.

Application of Standards 8.1 and 8.2 in the RETA Economies

Economy	A = Applied; P = Applied in Part; N = Not applied	
	Standard 8.1	Standard 8.2
Hong Kong, China	P	A
India	P	N
Indonesia	A	P
Japan	A	A
Korea	A	A
Malaysia	P	P
Pakistan	P	N
Philippines	A	P
Singapore	A	A
Taipei, China (Taiwan)	A	P
Thailand	A	P

Source: Asia Development Bank, "Report on Insolvency Law Reforms in the Asian and Pacific Region," *Law and Policy Reform Bulletin* 1: 10–86.

- Does it *apply* the standard?
- Does it *apply in part* the standard?
- Does it *not apply* the standard?

The evaluations were presented in tables and summarized for every country at the end of the report (see Table 8.4).

Following the categorical ratings of each country for each standard, the report provided summary observations, comments on the experiences of particular countries, and, sometimes, recommendations. While generally matter-of-fact in its descriptions, the evaluative commentary would sometimes further identify specific features in a country for criticism or praise. *Criticism* portrayed the practice in a country as a "matter of concern." The position in Thailand, Indonesia, and the Philippines was described as "not entirely satisfactory," or even said to "run [...] quite contrary" to the standard.²⁶ These critiques would sometimes be accompanied by a direct challenge to the country to remedy its fault—e.g., "Reform on this aspect is urgently required in Thailand."²⁷ General shortcomings were sometimes accompanied by general recommendations: e.g., "It is recommended that the absence, in some reorganization processes, of an initial automatic stay, should be remedied."²⁸

When handing out *praise*, the report might state that Indonesia has "a particularly good, low threshold, criteria";²⁹ or that Indonesian and Thai provisions on time limits "provide very good models";³⁰ or even that Hong Kong and Singapore

²⁶ Ibid., 47.

²⁷ Ibid., 33.

²⁸ Ibid., 36.

²⁹ Ibid., 32.

³⁰ Ibid., 38.

“provide the best examples.”³¹ Rhetorically, this performs an important function for it implies that remedies to problems can be found within the region. Moreover, it allows the ADB to positively reinforce countries which otherwise are criticized elsewhere. This careful allocation of praiseworthiness and criticism suggests the report has been crafted with care to induce change while remaining mindful of national sensitivities.

At the same time, the aggregation of results permits the ADB to rise above the nation-state. The report occasionally makes observations about the application of a standard throughout the region. For example, it may state: that a standard is generally applied throughout the RETA economies;³² that there is division among RETA economies by legal tradition;³³ that “only a limited number of RETA economies apply this standard”;³⁴ or that “there is a clear problem in the region in this area.”³⁵ Like the EBRD Legal Assessment Surveys, therefore, the indicators of compliance can be run in multiple directions: comparisons across standards in the same country; comparisons across countries on the same standard; grouped comparisons (e.g., by type of legal family); and regional generalizations. Unlike the EBRD legal indicator surveys, however, the one-shot ADB report does not permit comparisons across time within or among countries. Hence there is no capacity for the ADB to comment on trends towards or away from the good practice standards.

International Monetary Fund

The IMF Legal Department was also prodded into action following the Asian Financial Crisis. With the assistance of six distinguished insolvency experts from various parts of the world, the IMF produced a small book on *Orderly and Effective Insolvency Procedures* (hereafter *Insolvency*).³⁶ Significantly, the IMF deferred to the World Bank over the design of a systematic instrument of evaluation (see below). Instead, it produced a thoughtful document directed at two audiences—colleagues inside the IMF (see below), and policymakers in states where the IMF had intervened or might intervene in the future. For the latter, it sought to show (1) why insolvency reform is relevant to a policymaker, not only for resolving a crisis, but for promoting growth; (2) what issues must be addressed by any insolvency law or insolvency system; and (3) what key policy choices would follow, with their negative and positive consequences.

While the IMF book does not provide indicators or good practice standards, it was the only effort by the IFIs to provide a doctrinal and reasoned basis for universal

³¹ Asian Development Bank, *Law and Development at the Asian Development Bank* (n. 25); Asian Development Bank, “Report on Insolvency Law Reforms in the Asian and Pacific Region” (n. 25): 40.

³² Ibid., 29.

³³ Ibid., 28.

³⁴ Ibid., 31.

³⁵ International Monetary Fund, *Orderly and Effective Insolvency Procedures: Key Issues* (International Monetary Fund, 1999), 37.

³⁶ International Monetary Fund, *Orderly and Effective Insolvency Procedures: Key Issues* (n. 35).

Table 8.5 Principal conclusions on liquidation and distribution in bankruptcy

<i>Principal conclusions</i>
The procedure for liquidating assets should be timely and efficient and should provide for a sale that maximizes the value of the assets being liquidated. To that end, the law should normally allow for both public auctions and private sales, with the requirement that, in the latter case, the sale is either supervised by the court or approved by the creditors, or both. Adequate notice of any sale should be given to creditors.

Source: International Monetary Fund, "Orderly and Effective Insolvency Procedures: Key Issues" (Washington D.C.: International Monetary Fund, 1999).

norms and the only document to acknowledge variations in practice that would satisfy the same functions, i.e., the analog to a sociologist's structural equivalents of a functional imperative. In a sense, it lay behind the systematic Report on Observance of Standards and Codes (ROSC) developed by the World Bank and informed the subsequent successful effort at global norm-making by UNCITRAL.³⁷

Insolvency began by asserting that all legal systems, of whatever legal and historical provenance, must provide two frameworks: a *legal* framework and an *institutional* framework. The legal framework incorporates policy choices around nine issues in substantive and procedural law. The institutional frameworks include courts, insolvency professionals, and supporting professionals. For a major issue, whether legal or institutional, the IMF publication comes to a "principal conclusion" (Table 8.5). It was a short step for other IFIs to take the conclusions and turn them into indicators.

At the same time, the IMF book reveals a flexibility and contingency that is not so evident in some of the indicators developed by other IFIs, which take on a concrete and categorical sense that belies the varied contexts into which IFIs thrust global norms. Often *Insolvency* will acknowledge significant cross-national variation on a topic. In such cases, it sets up minimal criteria for insolvency systems and leaves it to national policymakers to determine how far they go beyond the threshold. For example, *Insolvency* indicates there should be a "stay" (a freeze on all actions by creditors to seize their assets from the debtor). However, given the different approaches taken by countries, the recommendation sets out minimal criteria for its operation anywhere.³⁸ In a close variant, *Insolvency* may propose a core minimum feature while indicating that additional features are desirable but not necessary. Hence "there must be [a] provision" on "x," but "countries should consider" . . . additional elements. Or "at a minimum," a list of transactions should be subject to avoidance (reversing recent transactions).³⁹ Later the UNCITRAL

³⁷ Halliday, Carruthers, and Block-Lieb, *Bankrupt: Global Lawmaking and Systemic Financial Crisis* (n. 4).

³⁸ IMF, "Orderly and Effective Insolvency Procedures" (n. 21), 30.

³⁹ *Ibid.*, 60, 37.

*Legislative Guide on Insolvency*⁴⁰ will follow a similar style of formal flexibility.⁴¹ *Insolvency* not infrequently expresses preferences but acknowledges the need for exceptions. For instance, the exclusion of any kind of enterprise from a country's insolvency law "should be avoided," but, in a gesture of tacit approval, "countries may wish" to exclude certain entities like financial institutions.⁴²

In short, while the IMF itself does not publish its own set of formal indicators, it simultaneously provided both the intellectual reasoning behind the choice of concepts on which indicators might be developed, and displayed a little more flexibility in its assessments than is often characteristic of formal survey or diagnostic instruments.

The World Bank

Carved into the marble of its main entry hall in Washington, DC, is the World Bank's aspiration to spearhead the cause of a "world free of poverty." Its geographical reach extends to all developing nations and its program commitments have therefore required long-term relationships with countries. The Bank also has an affiliated private investment arm, the International Finance Corporation (IFC), which may also have personnel located in countries where the IFC has substantial investments. All these initiatives are committed to building national infrastructure and institutions. The Bank's long-term programs are punctuated by involvement in emergency financial situations, in which it will usually partner with the IMF and regional banks in structural adjustment loans.⁴³ Commonly the Bank jump-starts reforms with adjustment loans and then follows them with long-term development projects.

As befits an institution committed to long-term developmental change, the Bank's Legal Department has been larger and more reform-oriented than any other IFI. During the 1990s, the Bank accepted the theory that sustainable development requires a comprehensive process that must also include legal and judicial reforms:⁴⁴ "Poverty cannot be fought and gains cannot be sustained without effective and equitable legal systems."⁴⁵ Given such logic, commercial lawmaking by the Legal Department might have been anticipated, but such was

⁴⁰ United Nations Commission on International Trade Law (UNCITRAL), *UNCITRAL Legislative Guide on Insolvency* (United Nations, 2004).

⁴¹ Susan Block-Lieb and Terrence C. Halliday, "Legitimacy and Global Lawmaking" in Fordham Law Legal Studies Research Paper No. 952492 (2006), SSRN.

⁴² IMF, "Orderly and Effective Insolvency Procedures" (n. 21), 21.

⁴³ "These are loans to finance general imports in the context of a program of policy reforms." World Bank, "World Development Report 2002: Building Institutions for Markets" (Oxford: Oxford University Press, 2002), p. 7; Interview 2022. See also Halliday and Carruthers, *Bankrupt* (n. 4), chs 5 and 6, on Indonesia and Korea during the Asian Financial Crisis.

⁴⁴ "Legal and judicial reform," states a publication of the World Bank "is one of the main pillars of the Comprehensive Development Framework that James D. Wolfensohn, president of the World Bank, proposed." World Bank Consultation Draft, "Effective Insolvency Systems: Principles and Guidelines," (World Bank, 2000).

⁴⁵ Ibid.

not the case. It took secular changes in economics and sharp financial shocks to invigorate the Bank's program towards commercial and insolvency lawmaking.

The accumulated experience of IFIs and the Bank in failed earlier phases of development led the Bank to the position that domestic and foreign private investment depends on the establishment of legal regimes that deliver stable and predictable justice. Indeed, the mantra of the rule of law⁴⁶ came to be accepted as an engine of economic development and effective governance. Law affects poverty: its absence, in the form of rampant corruption, "regressively taxes the poor." The benefits of globalization are distributed disproportionately in favor of the rich and deprive the poor when they have few legal rights or few realistic ways of pursuing them. This growing awareness of law's impact on markets, inequality, economic development, and governance came to a head with the sharp shocks to the international system since the late 1980s. The events that spurred the Bank to systematic commercial lawmaking began with the political and economic transformations of the former Soviet Union and the Comecon countries (members of the Council for Mutual Economic Assistance), and climaxed with the Asian Financial Crisis.⁴⁷ In the former, the Bank was called upon to aid transition countries to establish market institutions. In the latter, the Bank joined with the IMF in emergency actions to stem a regional crisis before it became global. Reforms of insolvency systems were injected into the agreements reached with Crisis countries.

The Legal Department at the Bank expanded its programs for legal and judicial reform significantly through the 1990s as it moved from an emphasis on disparate legal reform projects to judicial reforms and institution-building—a shift consistent with the Bank's growing commitment to matters of governance since 1991. In the decade from 1991 to 2002, the Legal Department funded some 330 projects. Indeed, from 1986 these involved 87 countries in 45 specialized areas of the law, ranging from property rights, utilities regulation, and privatization, to anti-corruption, judicial reform, and financial sector law.⁴⁸ The Legal Department has five principal financing instruments: (1) structural and sectoral adjustment loans which the Bank commonly extends with "conditionalities" for reform; (2) investment loans for institution-building and technical assistance that extend from two to five years; (3) Learning and Innovation Loans for small, fast, innovative projects; (4) the Adaptable Program Lending Loan for long-term programs in conditional phases; and (5) grants from its Institutional Development Fund. In addition, the Legal Department funds directly the advice given by its staff on draft legislation until such time as governments can finance consultants themselves.

⁴⁶ World Bank "World Development Report 2002 (n. 43): 3. According to the Bank "the rule of law prevails where (1) the government itself is bound by the law; (2) all in society are treated equally under the law; (3) the government authorities, including the judiciary, protect the human dignity of its citizens; and (4) justice is accessible for its citizens. The rule of law requires transparent legislation, fair laws, predictable enforcement, and accountable governments to maintain order, promote private sector growth, fight poverty, and have legitimacy."

⁴⁷ *Ibid.*, 1–3.

⁴⁸ *Ibid.*, 4, 9.

By the time the legal side of the Bank moved heavily into institutional reforms, the private financing and investment unit in the Bank was already seasoned in this area by the debt crises of the 1990s and their impact on corporate restructuring. In Ecuador, Mexico, and Venezuela, Bank officials outside the Legal Department began to develop models for corporate restructuring that could be implemented swiftly and forcefully in the event of a sharp contraction of the economy and widespread corporate insolvency. Rather than developing insolvency law, the corporate restructuring specialists advocated the creation of special administrative agencies that could quickly effect debt restructuring agreements between debtors and creditors.

The Asian Financial Crisis brought together the legal and corporate sides of the Bank in two mutually supporting ways. Because similar institutional weaknesses recurred in one after another of the Asian economies, the Legal Department heeded the call of the G-7 and G-22 for IFIs to pursue comprehensive legal and institutional responses to corporate failure in developing societies, both to stabilize crisis countries and to forestall future crises. The Legal Department appointed an American insolvency lawyer to lead the World Bank's institution-building initiative for insolvency regimes.

Assessments, or in-depth studies of national institutions, emerged strongly in the later 1990s as tools employed by the Bank to diagnose problems in laws and institutions, often as a prelude to reform programs.⁴⁹ The Bank's Legal Department created three interdependent assessment instruments in the field of corporate insolvency and creditor rights.

Principles and guidelines

The World Bank Legal Department was the last IFI to arrive on the global stage with its norms or instruments for national insolvency regimes. To differentiate itself from its sister IFIs, the Insolvency Unit within the Legal Department chose to create a set of master principles for all insolvency regimes across the world. Through these principles and linked assessment tools the Bank expected to provide resources for, and sometimes exert leverage over, policy- and lawmakers in developing countries.

The heavily funded Bank initiative proceeded by casting its consultative net very broadly indeed, not only to include previous IFIs and professional bodies that had undertaken norm-making on insolvency, but to reach out to scholars and practitioners across regions of the world.⁵⁰ The Insolvency Initiative constituted: an

⁴⁹ In the first five years, legal and judicial assessments were undertaken in Argentina (1995, 2001), Armenia (1998), Bulgaria (1999), Cambodia (2002), Ecuador (1994), Yugoslavia (2002), Georgia (1998), Indonesia (1998), Moldova (2002), Mongolia (2002), Mozambique (2002), Peru (1994), Romania (2002), Russia (2002), Sierra Leone (2002), Slovakia (2002), Trinidad and Tobago (1995), and Vietnam (2002). They continue to the present at the rate of about five a year. World Bank, "World Development Report 2002" (n. 43): 6–20.

⁵⁰ For instance, in its October 2000 Consultation Draft on the Principles, the Bank includes as "partner organizations" the African Development Bank, Asian Development Bank, European Bank for

Advisory Panel, which included the key insolvency specialist from each of the global multilateral, professional, and governance bodies interested engaged in insolvency reforms; a Task Force on Insolvency and Debtor–Creditor Rights, which included some of the world’s most distinguished practitioners, insolvency law reformers, and academic commentators; and no less than eight Working Groups of judges, practitioners, regulators, bankers, academics, and international civil servants.⁵¹ Each group was charged with preparing a report, preferably with recommendations, which the leader of the Initiative could draw upon in the drafting of the Principles and Guidelines for Effective Insolvency and Creditor Rights Systems (hereafter, the Principles). By any standard, these entities encompassed a global elite of the insolvency field, at least from advanced economies.⁵²

In contrast to the “standards” of the ADB or the “principal conclusions” of the IMF, the Bank documents (of which there were several drafts from 1999 to 2007) list some 30 “principles and guidelines.” In form, there is no discernible difference between the Bank’s principles and the ADB’s standards or the IMF’s conclusions. There is some difference in emphasis, however. As the text evolved it increasingly stressed the rights of creditors—a so-called legal framework for creditor rights. Unlike its sister IFIs, the final Principles placed a stronger emphasis on institutions than on substantive or procedural law.

Individual principles vary a good deal within the document. Some are general and aspirational without offering a point of view on any of the differences that divide the world’s insolvency systems.⁵³ A principle may mix a high-level aspirational statement (e.g., Principle 30 provides, “an insolvency system should be based on transparency and accountability”) with how it might be specified (“[R]ules should ensure ready access to court records”). There are principles that mix recommendations with justifications, such as Principle 16, on upholding *ex ante* rights negotiated by creditors in order to “encourage greater predictability in commercial relationships,” and Principle 23, which recommends that countries provide for a discharge “to ensure that the rehabilitated enterprise has the best chance of succeeding.” Some mix recommendations with non-specific criteria, such as Principle 13, which provides that asset disposal should be done “efficiently” and “at maximum values.” Others are precise and pre-statutory, with lists of issues

Reconstruction and Development (EBRD), Inter-American Development Bank, INSOL, the IBA, International Finance Corporation, IMF, Organization for Economic Cooperation and Development (OECD), and the United Nations Commission on International Trade Law (UNCITRAL). (“Effective Insolvency Systems: Principles and Guidelines”). *See also* “Principles and Guidelines for Effective Insolvency and Creditor Rights Systems,” (World Bank, 2001).

⁵¹ Working Groups on Institutional Capacity, Regulatory Framework, Business and Financial Sector, Debtor-Creditor Rights, Rehabilitation and Insolvency Alternatives, Economic Dimensions and Systemic Crises, State-Owned Enterprise Insolvency, and Bank Insolvencies.

⁵² Of the 86 people (with some duplication) listed in the Advisory Panel, Task Force and Working Group, approximately eight were from developing countries. None were from transitional countries.

⁵³ *See* Principle 1 (“A modern credit-based economy requires predictable, transparent and affordable enforcement of both unsecured and secured credit claims . . .”) and Principle 18 (“The law should provide for a commercially sound form of priority funding . . .”).

to cover.⁵⁴ While most indicate a single preference, there are one or two principles that articulate a preference between express alternatives (e.g., Principle 9 C, which lists the “preferred test” as the “liquidity test,” but accepts a balance sheet test alongside it as “secondary”). Rarely, two alternatives are given in the principle itself without any indication of the Bank’s preference (cf. Principle 11 on Governance: Management where “two preferred approaches” are given equal weight).

For the Bank’s Legal Department, therefore, the Principles specify the normative parameters against which indicators will be developed and assessments undertaken.

The Assessment Template

The Insolvency Initiative gave the Principles deeper bite by developing two more penetrating instruments and practices: the design of an Assessment Template that enables countries to engage in self-assessment of compliance with the Bank’s norms, and the practice of national evaluations or reviews by the Bank. By either method the Bank could drive countries towards active engagement with the Principles and thereby ensure the norms did not sit on a shelf awaiting the next emergency.

The Template keyed into the Principles with a set of practical questions that stimulated national officials to appraise for themselves whether each of the 30+ principles was functioning in practice. This exercise at self-diagnosis could prompt quite precise evaluations. On creating a security interest, the Template wanted to know how much it cost and how long it took. On the liability of company directors and officers, the Template asked how often penalties for insolvency-related crimes were applied, and how often cases were successfully brought against directors and by whom. For the insolvency system as a whole the Template expected that countries would have in place “continuous monitoring.”

The questions themselves effectively constitute checklists of topics that any policymaker should consider in the reform of substantive bankruptcy law. Take the area of creditor rights. Under Principle 2, the Template presents 15 questions relevant to “Judicial Remedies” (Table 8.6). The heading itself clearly indicates a normative assertion that there should be judicial remedies, and the questions equally clearly flag their normative scope.

Likewise, under Principle 3 on security interest legislation, the Template asks, “Which of the following does the law allow to be used as collateral?” and then provides a list of all known forms of collateral (e.g., immovable and movable property, tangible and intangible property, equipment, inventory, accounts receivable, future property, on a shifting pool of assets, based on possessory and non-possessory interests, etc.).

The connection between the Principles and the Template signifies a move that illustrates the often fraught relations between normative concepts and their oper-

⁵⁴ See, for instance, Principle 12 (“... The law should provide for such things as general creditors assembly for major decisions, to appoint the creditors committee and to determine the committee’s membership, quorum and voting rights, powers and the conduct of meetings”).

Table 8.6 Template questions on judicial remedies, insolvency ROSC

-
- Describe the procedure for seizing immovable property to satisfy an unsecured right.
 - Describe the procedure for selling immovable property
 - Describe the procedure for disbursement of proceeds from sale of immovable property.
 - Describe the procedure for seizing movable property
 - Describe the procedure for selling movable property
 - Describe the procedure for collecting on intangible assets
-

ationalization. The Principles are sensitive to criticisms of the World Bank in general, and early drafts of the Principles, in particular, which claimed that they push homogeneous “one-size-fits-all” norms on a heterogeneous world, a homogeneity of normative expectations that reflect too strongly the powerful pressures of the United States on Bank policies and practices. In response to criticisms that the first drafts of the Principles seemed to export US law and institutions across the world, later drafts hedged such imperatives with a tolerance of alternatives. As a result, the final version of the Principles appears more flexible in orientation and functional in expectations.

Whereas the Principles represent the public face of the Bank’s norms, the Template is relatively invisible to critics, scholars, and commentators. There is a hint in the Template that its invisibility gave license to Americanists in the Bank to tone down their “tolerance” of the diversity apparent in the Principles and to emphasize, instead, the preferences of US lawyers in the Template.

For instance, one of the most controversial differences in world insolvency systems relates to management of companies during rehabilitation: should current management stay in place, under supervision by creditors, administrators, or courts; or should management be displaced by an independent administrator? Principle 11 allows both, which is a shift from an earlier preference for debtors. But the Template revisits this even-handedness and hints at a preference for the US practice of debtors/management in control. Whereas Principle 8 asserts that law should offer procedures for liquidation and rehabilitation, the Template unexpectedly raises the questions: “Does the law provide for the concept of ‘debtor in possession’, where the debtor continues to operate the enterprise on a day-to-day basis, but under close supervision of the administrator? Which powers does the ‘debtor-in-possession’ enjoy?” The questions clearly suggest that a law should have a provision for managers to stay in control of their companies. And the placement of this suggestion, several principles before an alternative is given, reinforces the point. Later, at Principle 11, when the two alternative models are recognized, the first question under “Appointment of an Administrator” asks: “Does the law provide for the debtor to keep estate management?” It follows this with the coining of a new term: “debtor-in-management,” which breaks away from the American term, “debtor-in-possession,” but nonetheless retains its meaning. In each case, the Template makes clear that management requires supervision. And a question equally clearly raises the alternative: “Does the law provide for the alternative appointment of a bankruptcy administrator?” Yet a tilt remains towards debtor-

in-possession or debtor-in-management. A similar analysis can be applied to the preference implicit in whether or not a country has a specialized insolvency court.⁵⁵

Even more than the Principles, the Template underscores the importance of institutions by asking questions that probe unsubtly into their structures and operations. On registries for securities, for instance, the Template wants to know:

- What body is responsible for operating and maintaining the registry?
- What is the responsibility of the registering body?
- Where is the registry located?
- Is the registry electronic or manual?
- How often is the registry updated?
- Are there any transparency requirements?
- Do registry officials review filings for accuracy and legality?
- What are the procedures involved in adding an interest to the registry?
- Who has access?
- Do multiple registration systems exist? (If yes, how are they coordinated?)

The implicit norms are plain: there should be a clear, coordinated system of registries throughout a country whose locations are well-known. Preferably, the registry should permit easy electronic access. There should be a hierarchical structure of responsibility and an internal review process. These implicit norms conveniently overlook the fact that systems of registries are highly controversial in certain countries, most especially Germany and some other civil law countries. Under cover of operationalization or specification of the principles, therefore, comes a shift in normative emphasis and a push towards harmonization on a US model.

Reports on the Observance of Standards and Codes (ROSCs)

In the aftermath of the Asian Financial Crisis, the global IFIs initiated a comprehensive effort to evaluate the conformity of countries with global standards that relate to international financial stability. According to the World Bank:

The international community has emphasized the important role of international standards in strengthening the international financial architecture. In a world of integrated capital markets, financial crises in individual countries can imperil international financial stability. This provides a basic public goods rationale for minimum international standards which would benefit both international and individual national systems. At the international level, international standards enhance transparency as well as multilateral surveillance. They help to better identify weaknesses that may contribute to economic and financial vulnerability, foster market efficiency and discipline, and ultimately contribute to a global economy which is more robust and less prone to crisis. At the national level, international standards provide a benchmark that can help identify vulnerabilities as well as guide policy reform. To best serve these objectives, however, the scope and application of such standards needs to be assessed in

⁵⁵ "Principles for Effective Insolvency and Creditor Rights Systems," World Bank (2005), Principle 27, p. 51.

the context of a country's overall development strategy and tailored to individual country circumstances.

In this connection, the IMF has invited the World Bank to embark on a joint pilot exercise preparing Reports of the Observance of Standards and Codes (ROSCs). In this exercise, the two institutions are undertaking a large number of summary assessments of the observance of selected standards relevant to private and financial sector development and stability. These assessments are being collected as "modules" in country binders constituting the ROSCs. Under this modular approach, the Fund takes the lead in preparing modules in the area of data dissemination and fiscal transparency. Modules for the financial sector (monetary and financial policy transparency, banking supervision, securities market regulation, payment systems, deposit insurance) are mostly derived as by-products from a parallel Bank-Fund Financial Sector Assessment Program (FSAP). The World Bank has been asked to take the lead in three areas covered by ROSCs: (i) corporate governance, (ii) accounting and auditing, and (iii) insolvency regimes and creditor rights.⁵⁶

Coincident with the development of the World Bank Principles and Guidelines for Effective Insolvency and Creditors Rights Systems, the Bank Legal Department designed a ROSC for insolvency systems. Unlike the Template, which is a self-assessment by a state, the ROSC on insolvency was a Bank-led staff appraisal of a country's insolvency system. The insolvency ROSCs had three elements—design, data collection, and reports. The data collection proceeds through two main stages: (1) "an agreement among the country team in Washington, the World Bank official on-site in the country, and the country itself," preceded by an analysis of extant substantive law; and (2) a two-week country mission, during which Bank staff

interview up to fifty or sixty people, including public and private bankers, government officials (e.g., Central Bank, Ministry of Justice, Ministry of Finance), lawyers (usually those that are in-country counsel to the World Bank, a couple of international law firms, and a couple of local firms), accountants (if possible, a couple of Big Four firms), sometimes officials from debtor organizations (e.g., Chambers of Commerce, representatives of industry associations), judges, and officials in government restructuring agencies.⁵⁷

The World Bank team returns to Washington and prepares a report which it delivers to its government counterparts. Governments have the option to publish nothing, to publish the Executive Summary only, or to publish both the Executive Summary and the full report.⁵⁸ Most elect to publish nothing or the Executive Summary only.

Until 2005 the ROSC assessments were purely qualitative—a long series of written responses to questions or summary observations. After 2005, and an internal IMF/World Bank evaluation of the ROSC program, there was a move to ordinal ranking of countries, principle-by-principle, on a four-point scale (see

⁵⁶ From <http://www.worldbank.org/ifa/rosc_more.html> (accessed August 10, 2010).

⁵⁷ Interview 2756.

⁵⁸ Postings of ROSCs can be found at <<http://go.worldbank.org/XWJZ8NZSHO>> (last accessed April 12, 2012). Only six national reports were posted.

Table 8.7 ROSC rankings, World Bank, 2010

Observed (O) whenever all essential requirements of the Principle are generally met without any significant deficiencies;

Largely Observed (LO) whenever only minor shortcomings impede full observance of the Principle;

Materially Non-observed (MNO) whenever the legislation or practices derogate significantly from the Principle; and

Non-observed (NO) whenever there is no relevant legislation or whenever the latter or the practices do not fit in at all with the Principle.

Source: "The Standards and Codes Initiative—Is It Effective? And How Can It Be Approved?" World Bank, International Monetary Fund (2005).

Table 8.7).⁵⁹ The rankings were then collated in a summary table at the beginning of the ROSC Report. Significantly, these results were disclosed privately to country officials, *not* to the public. Moreover, it is obvious that it would be an easy step to convert the rankings into numerical values and to engage in various arithmetical and statistical analysis. The World Bank Legal Department insists it does not do so itself, but it is likely that other consumers of the ROSCs do so for their own purposes (see below).

Politics and theory of yardsticks

Although the timing varies, at one moment or another over the past decade each of four IFI Legal Departments has measured law using indicators consistent with the criteria enumerated by Davis et al.⁶⁰ (Table 8.8). The EBRD fully satisfies all criteria and the ADB comes close, as do the IMF and World Bank since 2005.

However, to refine the overview of legal diagnoses a little more, the insolvency diagnostic tools developed by the four IFIs might more appropriately be divided into indicators and pre-indicators (Table 8.9). Those of the EBRD and ADB permit comparability over topics and countries. The EBRD adds the possibility of comparability over time so progress, or its absence, can be weighed and reported. By contrast, only the IMF provides clear policy options and a considered juridical and theoretical framework. The IMF and World Bank ROSCs until the mid-2000s did not permit the possibility of ordinal ranking, but they did permit other organizations to take the intelligence gathered systematically in the ROSCs and translate those judgments into such rankings. To comprehend this key move requires a broader context, one of politics within and among IFIs. These help explain the push or pull of indicators as a diagnostic device by the legal departments.

⁵⁹ "The Standards and Codes Initiative—Is it Effective? And How Can It Be Approved?" World Bank, International Monetary Fund (2005).

⁶⁰ Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 2), 73–8.

Table 8.8 IFI legal department instruments classified by author according to criteria enumerated in Davis et al., “Indicators as a Technology of Global Governance”

	EBRD	ADB	IMF/WB ROSCs ^a
Ordinal	Y	Y	Y
Simplification	Y	Y	Y
Expertise	Y	Y	Y
Social science methodology	Y		
Inter-unit comparability	Y	Y	Y

^a After 2005.

Source: T. Halliday, 2012.

Table 8.9 Uses of FI pre-indicators and indicators

	Indicators		Pre-indicators (to 2005)		Indicators (since 2005)
	EBRD	ADB	IMF	World Bank/IMF ROSCs	World Bank/IMF ROSCs
Policy choices			X		
Cross-topic comparisons	X	X		X	X
Cross-time comparisons	X				
Cross-country comparisons	X	X			
Public ratings	X	X			

Source: T. Halliday, 2012.

The ecology of IFIs

It is useful theoretically to imagine the array of IFIs in any legal domain as organizations which exist in an ecology.⁶¹ A century-long school of sociological scholarship has elaborated a rich armory of concepts and theory to account not only for community studies, where ecological theory originated, but more recently for organizational studies, where ecological theory thrives.⁶² The ecology itself is situated within political, economic, and social niches which provide it with

⁶¹ See Susan Block-Lieb and Terence C. Halliday, “Social Ecology, Recursivity and Temporality: A Sociology of Global Law-Making” in *American Sociological Association Annual Meeting*, 2011.

⁶² Andrew Abbott, “Linked Ecologies: States and Universities as Environments for Professions,” *Sociological Theory* 23 (2005): 245–74; Andrew Abbott, *Ecologies and Fields* (University of Chicago, 2009); John H. Freeman and Pino G. Audia, “Community Ecology and the Sociology of Organizations,” *Annual Review of Sociology* 32 (2006): 145–69; Michael T. Hannan, “Ecologies of Organizations: Diversity and Identity,” *Journal of Economic Perspectives* 19 (2005): 51–70; Richard W. Scott, “Reflections on a Half-Century of Organizational Sociology,” *Annual Review of Sociology* 30 (2004): 1–21.

resources.⁶³ The international organizations (IOs) within the ecology compete or cooperate or bargain over the resources available to them and the power relations that define their interactions. Survival or ascendancy in an ecology requires that an IO obtain sufficient resources, or win competitive struggles, or manage a division of labor such that an IO can reproduce itself in changing circumstances both inside and outside the ecology.

Seen in these terms, the IFIs in the area of corporate insolvency law are subject to contingencies outside the ecology and within it. The emergence of the corporate insolvency ecology itself can be traced back to three sets of developments. First, in the 1980s, a growing number of large corporate failures in the United States and Britain focused the attention of experts, especially professionals, on the need for domestic systemic solutions.⁶⁴ When business failures occurred in massive multinational corporations, the problem became even more acute because no nations and no transnational protocols existed to liquidate or reorganize failing multinationals in an orderly way. This produced a double pressure on advanced economies—to reform their domestic laws and to find transnational mechanisms for orderly bankruptcy proceedings. Secondly, the geopolitical reconfiguration of Central and Eastern Europe compelled IFIs, together with national and international professional associations, to produce norms and mechanisms for their implementation and evaluation. The sudden transformation simultaneously of economies in 20–30 countries compelled the EBRD in particular to invent ways of appraising quickly the pace and scope of institution-building in law and markets and to monitor whether those countries were consolidating durable institutions. Thirdly, the Asian Financial Crisis propelled regional (ADB) and global development banks (IMF, World Bank, IFC) to undertake rapid diagnoses and find quick prescriptions to stabilize the regional economies and forestall a recurrence or domino effect elsewhere.

The ecology that bred indicators, therefore, came into being through circumstances that were national, geo-political, political-economic, and financial—all of which were mediated by professionals and refracted through their frames of professional interests and epistemologies (see below). The political, economic, and financial environs concomitantly “selected” certain IOs (e.g., through the direction of the G-7 and G-22) to create an ecology and stimulated others to enter it (e.g., OECD, UNCITRAL). By 1998, several IFIs, two major international professional associations (International Bar Association, the International Federation of Insolvency Practitioners), international governance bodies (UNCITRAL, OECD), and certain sovereign states (most notably the United States), all inhabited the ecology.

Like any organizations in a given ecology, the IOs and other actors found themselves potentially in competition for resources and ascendancy in the

⁶³ Pamela A. Popielarz and Zachary P. Neal, “The Niche as a Theoretical Tool,” *Annual Review of Sociology* 33 (2007): 65–84.

⁶⁴ Bruce G. Carruthers and Terence C. Halliday, *Rescuing Business: The Making of Corporate Bankruptcy Law in England and the United States* (Oxford University Press, 1998).

development of the “gold standard” of global norms for corporate insolvency law. Their incipient and later open competition proceeded through one or another set of norms.⁶⁵ Either they produced alternative sets of codified norms and/or they created diagnostic instruments which were sometimes derived from formal norms or were implicitly normative themselves.⁶⁶ Potentially juggling for ascendancy were the ADB’s standards, the EBRD’s legal transition surveys, the IMF’s recommendations, and the World Bank’s principles. To complicate matters even further, UNCITRAL entered the fray in 1999, potentially to eclipse or transcend all the others.

Sometimes it is possible for organizations in an ecology to negotiate a settlement which allows all of them to survive as players. This is precisely what happened in the corporate insolvency ecology. The ADB and IMF fed their norm-making efforts into UNCITRAL. Initially, it seemed the World Bank would do the same, but it broke out into a sometimes fraught competition with UNCITRAL, so much so that the US State Department, the General Counsel of the UN, and the General Counsels of the IMF and World Bank had to intervene, under the oversight of the US Treasury, and work out a concordat.

The Solomonic solution essentially resulted in a division of labor: UNCITRAL’s *Legislative Guide on Insolvency*⁶⁷ would be acknowledged as the principal body of global norms, complemented by the much briefer World Bank Principles.⁶⁸ The worldwide diagnostic functions of the ecology, however, were delegated to the World Bank via its insolvency ROSC. In short, a political settlement of competing IFIs and IOs in the international trade law ecology was facilitated by a division of labor where one IO (UNCITRAL) could claim ascendancy in its detailed global norm-making, while another IO (World Bank) would take responsibility for diagnosis of legal systems that would lead countries to conformity with UNCITRAL norms. The distinctiveness of the World’s Bank value-added came initially in its systematic representation of the UNCITRAL norms in long diagnostic instruments and later in its conversion of those qualitative diagnoses into ordinal rankings of compliance. Thus, a shift towards indicators by the World Bank both justified the distinctive role of the Bank Legal Vice-Presidency in diffusion of trade law norms and ratcheted up the pressure on nation-states to adopt UNCITRAL norms. Hence the division of labor cohered into an inter-dependency of norm-making and norm implementation, and arguably an enhancement of the power and influence of each IO. Diagnostic instruments coexisted inter-dependently with codified norms.

⁶⁵ Terence C. Halliday, Susan Block-Lieb, and Bruce G. Carruthers, “Attaining the Global Standard” in Terence C. Halliday and Bruce G. Carruthers (eds), *Bankrupt: Global Lawmaking and Systemic Financial Crisis* (Stanford University Press, 2009), 122–65.

⁶⁶ Terence C. Halliday and Bruce G. Carruthers, “The Recursivity of Law: Global Normmaking and National Lawmaking in the Globalization of Bankruptcy Regimes,” *American Journal of Sociology* 112 (2007): 1135–202.

⁶⁷ United Nations Commission on International Trade Law (UNCITRAL), *UNCITRAL Legislative Guide on Insolvency* (United Nations, 2004).

⁶⁸ World Bank, “Principles for Effective Insolvency and Creditor Rights System” (World Bank, 2005).

Organizational micro-politics

If the politics within the corporate insolvency ecology produced a settlement among potentially competitive IOs, the dynamics of settlement within the ecology were driven in substantial part by politics inside IFIs.

Much of these politics turns on the historically marginal but currently more central situation of IFI legal departments within their own organizations. Although the story is too complicated to tell here, we may style the transformation of IFI legal departments over the past two decades as a form of status mobility.⁶⁹ Traditionally assigned a role roughly comparable to house counsel in corporations, the ideational shift in development theory towards the importance of institutions, with law prominent among them, began to propel in-house IFI lawyers away from the implementation tasks of contract writing and regulatory oversight towards law reform as a bulwark of macroeconomic development. Organizational politics manifested themselves in several ways.

First, before the 1990s IFI legal departments had not been required to undertake diagnostic tasks. Their role had been to service the financial engine of the development banks. The law and development turn within the IFIs now required them to take on new tasks—and tasks for which most were not particularly well equipped. Whereas the economists and finance specialists in the IFIs had sophisticated technologies of macroeconomic evaluation, the legal departments had none. The urgent need for institution-building was immediately manifest to the early legal officers of the EBRD given the absence of commercial law, courts, regulatory institutions, qualified professions—indeed, of the entire panoply of legal market infrastructures—in the former command economies. The EBRD Legal Department began its life with law reform on its agenda. Law reform required both diagnosis and prescription. The development of diagnostic instruments, and the push to institutionalize reforms in fragile states and markets, pressed EBRD lawyers to find quick and easy methods to assay institutional vacuums, provide remedial advice, and monitor actual implementation. Since governments did not collect the kinds of systematic statistics in commercial law that were readily available to economists in the financial sector, the EBRD chose simple reputational mail surveys as its first expedient. Essentially it opted for a less labor-intensive method to gather data from all its transition economy members. The ADB, by contrast, chose a more labor-intensive method of country-by-country evaluations by a consultant. However, it could not possibly cover all developing nations in its region and therefore restricted itself to 11 countries.

⁶⁹ For sociologists, collective status mobility refers to the movement of an occupation or other collectivity up or down a status hierarchy. This may occur within a society, an organization, or some other embedding structure, such as an international financial institution. Cf. John P. Heinz and Edward O. Laumann, *Chicago Lawyers: The Social Structure of the Bar* (New York: Russell Sage, 1982); Magali S. Larson, *The Rise of Professionalism* (Berkeley: University of California Press, 1977).

Secondly, the demands of other departments within IFIs intensified the push towards indicators. The EBRD specializes in loans to the private sector in transitional economies. Like any good bank, it evaluates the quality of the applicant. Its evaluation relies in substantial part on risk assessment models. Since legal risk is now an element of overall risk assessment, an ability to insert numbers into risk models is a benefit to loan officers. Hence there was a pull from the financial side of the EBRD to obtain systematic data that might permit legal indicators to be inserted into risk assessment models.

Initially the designers and implementers of IMF/World Bank ROSCs did not provide systematic data to sections of the Bank or Fund that were approving its various loans and forms of technical assistance. However, a 2005 internal evaluation of ROSCs found that their impact was insufficient for three audiences, all of whom noted “shortcomings of the product.”⁷⁰ Inside the Bank, mission chiefs and others said that “the significance of ROSC findings is not necessarily clear to non-specialists.”⁷¹ It was not obvious which factors contributed to financial vulnerability or promoted stability. Outside the Bank, “market participants”—rating agencies, private banks—used ROSC information very little in their investment decisions, not because it was irrelevant but because its format did not easily permit them to deploy it in their decision-making. According to the internal evaluation, even civil society groups complained over the technical language of the documents, noting that ROSC findings were “not easy to decipher.”⁷²

Pressure arose from three sides, urging the Fund and Bank to provide some combination of (a) executive summaries; (b) a summary table which ranked each principal criterion in a review on a four-point scale; (c) a prioritized set of key recommendations; and (d) the development of “quantitative measures of compliance that could be included in risk models.”⁷³ Armed with systematic data in this form, the Fund and Bank might then engage in “cross-country” and “inter-temporal comparisons.”⁷⁴ In short, although the evaluation method differed, the IMF/World Bank internal evaluation recommended that ROSCs be reported in a manner more consistent with the EBRD’s transition surveys.⁷⁵ This move would produce packets of indicators that were ordinal, simple, reflective of expertise, and permitted inter-unit comparability.⁷⁶

Third, the pressures on legal departments to use and report indicators result in part from variations in the forms of leverage the IFIs exert over their client countries and private borrowers. Opportunities for the use of explicit conditionality come rarely to the Fund and the Bank, in part due to the backlash against the deployment

⁷⁰ IMF, “The Standards and Codes Initiative—Is it Effective? And How Can It Be Approved?” (World Bank, International Monetary Fund 2005), paras 12, 41.

⁷¹ *Ibid.*, para. 34, p. 23.

⁷² *Ibid.*, para 29, p. 48.

⁷³ *Ibid.*, paras 14, p. 42; 15, p. 31; Table 8.7, p. 32.

⁷⁴ *Ibid.*, p. 33.

⁷⁵ Ironically, this recommendation came at much the same time as the EBRD Legal Department was backing away from opinion surveys and quantitative reports.

⁷⁶ Davis, Kingsbury, and Merry, “Indicators as a Technology of Global Governance” (n. 2), 73–8.

of economic coercion as a condition of structural reforms. Signals of IFI displeasure can become rather blurred in the granting or withholding of loan support and technical assistance. As a result, the regional development banks—EBRD and ADB—rely rather more on public shaming to press countries into reforms. In both cases, the European and Asian IFIs combined verbal affirmation or critique of a country's compliance with stark, cross-national, public report cards on the relative performance of states vis-à-vis each other. Public displays of a nation's low scores in comparison to its neighbors, it was thought, focus the attention of national policy-makers on reforms not least because those comparisons may also focus the minds of foreign investors. Use of indicators and tables of results provide a parsimonious and efficient way to exercise the power of persuasion.

Professional epistemologies and competition

There is a politics of professionalism within IFIs that also stimulates recourse to indicators. In Andrew Abbott's terms,⁷⁷ IFIs are a site for professional competition, both in the diagnosis and prescription of reforms. Since IFIs are financial institutions, they are the natural territory of economists and finance specialists. Their principal concerns are macroeconomic. Their charges are to reduce poverty and facilitate economic growth. Their principal inhabitants, therefore, are professionals who proffer expertise in markets.

Since the institutional turn in economics, however, law and legal institutions have provided a wedge for lawyers. Economists have come to believe that substantive law and legal institutions (courts, regulatory agencies, out-of-court mediation bodies) are necessary structures to enable and constrain markets. Hence a convergence has occurred between new institutional economics and long-standing jurisprudential traditions in legal realism, law and development, and law and society. Moreover, the very opacity of black letter law, its complexity and technicality, about which the IMF/WB internal report on ROSCs complained, gives IFI lawyers a distinctive technical authority which they can wield to their advantage. Exercising policy influence under the guise of technical expertise is a tactic of long standing among legal professionals.⁷⁸

Yet they cannot wield their unique authority with complete disregard of its limitations. One limitation relates to institution-building itself, a subject that is

⁷⁷ Andrew Abbott, "Jurisdictional Conflicts: A New Approach to the Development of Legal Professions," *American Bar Foundation Research Journal* (1986): 187–224; Andrew Abbott, *The System of Professions: An Essay on the Division of Expert Labor* (Chicago, IL: University of Chicago Press, 1988).

⁷⁸ Terence C. Halliday, "Knowledge Mandates: Collective Influence by Scientific, Normative and Syncretic Professions," *British Journal of Sociology* XXXVI (1985): 421–47; Terence C. Halliday, *Beyond Monopoly: Lawyers, State Crises, and Professional Empowerment* (Chicago, IL: University of Chicago Press, 1987); Andrew Harding and Amanda Whiting, "Custodial of Civil Liberties and Justice in Malaysia," in Terence C. Halliday, Lucien Karpik, and Malcolm M. Freeley (eds), *Fates of Political Liberalism in the British Post-Colony: The Politics of the Legal Complex* (Cambridge University Press, 2012).

usually absent from the education of lawyers. Another limitation is the audience for lawyers: they address not only legal specialists but also non-specialists in other professions. Thus lawyers in IFIs must find the delicate balance between the wielding of their unique technical authority and its attendant increments of power while at the same time tailoring its diagnoses persuasively to non-technical audiences.

Their optimal strategy for influence therefore combines technical complexity with diagnostic accessibility. The former protects their professional jurisdiction from intrusion by other professions within the IFI, most notably the economists and finance specialists. It gives them a mandate to build authority-buttressing ties to legal specialists in sister IFIs and other global bodies such as bar associations and global governance organizations (e.g., OECD, UNCITRAL). Not least, technical complexity both protects and authorizes legal departments to be direct interlocutors with national policymakers whose institution-building in markets and elsewhere almost always proceeds through legal instruments, whether forms of law or types of legal institutions.

The latter—non-technical accessibility—multiplies the leverage of lawyers inside their respective IFIs, since they now can provide meaning in packages consistent with other types of quantification used to evaluate countries and assay risk. Indicators likewise enhance the legitimacy of the IFI to outside bodies whose support is critical. Private investors and banks are the most notable of these, but increasingly IFIs also look to civil society groups who may have much less expertise than the financial industry to cull through long pages of technical legal evaluations.

The increasingly ambitious interventions of IFIs into cognate institutions to markets—political stability, government effectiveness, corporate governance, regulatory quality, rule of law, control of corruption—further strengthen the hand of law departments inside IFIs. Arguably lawyers have a greater capacity to convert the policy or moral judgments associated with these elements of state-building into technical specifications and prescriptions than their professional competitors.⁷⁹ If the expansiveness of IFI agendas broadens the mandates of the IFIs, then the adeptness of legal departments in converting policy or moral issues into technical matters helps insulate IFIs from the perennial criticisms that they are intervening in domestic politics and far exceeding their macroeconomic bounds. In this respect, indicators conveyed back to national capitals through IMF/World Bank ROSCs drive national policymakers into the arms of the IFIs lawyers whose technical assistance will be essential to remedy the faults supposedly highlighted by low scores in the indicator tables.

⁷⁹ Terence C. Halliday, "Knowledge Mandates: Collective Influence by Scientific, Normative and Syncretic Professions" (n. 78), 421–47.

Legitimation

Ecological theories of organizations emphasize the recourse to a politics of legitimation by any organization seeking to enter, consolidate, or compete within the ecology.⁸⁰ We observe three manifestations of legitimation politics in the globalization of norms for trade law: among IOs and IFIs within the ecology; within each of the organizations in the ecology; and between the IOs in the ecology and the nation-states they seek to influence.

First, the competition to be accepted as a legitimate organization in a global norm-making ecology requires that each IO justify—to itself, its audiences, and other IOs in the ecology—that it belongs. IOs actively seek concomitantly to buttress their own claims to be legitimate builders of norms and to de-legitimate competitors. In assessing their legitimacy, in the corporate insolvency arena, IOs variously relied on strong claims to expert consultation, to alliances with the dominant professional associations, to representativeness of norm-makers from the main types of economies (developing, transitional, advanced), and to incorporation of principal economic interests in deliberation.⁸¹ At the same time they not so subtly signaled why their claims to recognition as norm-makers and/or norm-diagnostics had stronger warrants than their competitors. Indicators—the visible expression of diagnostic capabilities—sharpen some legitimation claims vis-à-vis other actors in the ecology.

Secondly, since the internal politics of IFIs manifest themselves sometimes as struggles among departments within the IFI and sometimes as internal organizational contestations of professional epistemologies and power, a complex politics of legitimation comes into play on both counts. While it is the case that recourse to putatively scientific instruments may buttress the authority of legal departments and the diagnoses of legal institutions in relation to other sections of an IFI, it is also the case that legal departments can respond to indicators developed elsewhere in their institutions with skepticism. For instance, for a variety of reasons not yet fully evident, the World Bank Legal Vice-Presidency became quite skeptical of the indicators and inferences of the *Doing Business* survey produced elsewhere in the Bank. Those reservations are echoed by other IFIs: some IFI lawyers would like nothing more than to see the *Doing Business* steamroller break down, or at least spend more time in the repair shop.⁸² Hence indicators become a battleground

⁸⁰ Thomas M. Franck, "The Power of Legitimacy and the Legitimacy of Power: International Law in an Age of Power Disequilibrium," *American Journal of International Law* 100 (2006): 88–106; John H. Freeman and Pino G. Audia, "Community Ecology and the Sociology of Organizations" (n. 62), 145–69; Richard W. Scott, "Reflections on a Half-Century of Organizational Sociology," *Annual Review of Sociology* 30 (2004): 1–21; Mark C. Suchman, "Managing Legitimacy: Strategic and Institutional Approaches," *Academy of Management Review* 20 (1995): 571–610.

⁸¹ Susan Block-Lieb and Terence C. Halliday, "Legitimacy and Global Lawmaking," in Fordham Law Legal Studies Research Paper No. 952492: SSRN.

⁸² See Proceedings of Center of Law and Globalization Roundtable with Legal Departments of International Financial Institutions, "Measuring Law: How to do it right in the real world," Washington, DC, March 23, 2010.

within IFIs over theory and methods, prescriptions and leverage—battles that influence the relative power of organizational groups, their access to resources, and their privileged status in IFI decision-making.

Thirdly, the normative impact of IFIs depends substantially on the legitimacy they and their technologies are accorded by national policymakers, technocrats, and reformers. Legitimation inheres not only in the methods by which global norms are developed, but in the formal products of the IOs themselves.⁸³ UNCITRAL claims ascendancy for its norms both because they emerged from an international deliberative process and because its 500-page text offers a comprehensive, thoughtful, reasoned presentation arrayed from abstract governing principles down to specific statutory language. In the same way but in different forms, the IMF crafted its 90-page discussion of policy options for lawmakers, and the World Bank offered its parsimonious 33 principles, which summarized much of the content of both UNCITRAL and the IMF at a higher order of abstraction.

If states are to conform themselves to global norms—and national leaders will frequently make this a matter of expediency—then they need fairly clear yardsticks of where they fall short, what must be remedied, and how. This global-local engagement, however, is a matter of great delicacy. What legitimates IFIs vis-à-vis each other may delegitimize states. Indicators are controversial. They grab the attention of authorities and provide practical guidance for programmatic adaptations; yet they also potentially embarrass nations before their peers and discredit elites responsible for institutions under evaluation. In fact, the divergent interests of IFIs in pressuring states, national elites in preserving their reputations, and state reformers in breaking domestic political impasses, produce a very complicated politics of legitimation.

IFI legal departments deploy their indicators in two broad ways: public display and private communication. Public display, and often shaming, is best exemplified by *Doing Business* reports, but is clearly among the aims of the ADB's 2000 report and EBRD reports to the present. Private communication of rankings is the preferred method of the IMF and World Bank, at least on insolvency rankings through ROSCs. But indicators are also potent potential weapons in the hands of domestic reformers who confront resistance from national policymakers. While some IFIs may spare current elites shaming by quietly conveying rankings to them, they thereby also potentially deprive other parties to reform of the fuel they need to mobilize domestic proponents of change.

Recursivity

If the use of indicators by IFI legal departments must be understood in relation to the organizational politics within each IFI and the ecological politics among IFIs, there is another and more structured set of relations which link the development of indicators to legal change.

⁸³ Halliday, Block-Lieb, and Carruthers, "Rhetorical Legitimation: Global Scripts as Strategic Devices of International Organizations" (n. 5), 1–36.

Legal change in a global context can be theoretically understood as a recursive process.⁸⁴ Major reform movements, including those in global trade, proceed through episodes which start with facilitating conditions and then precipitating events, and conclude with “settlements” at a new equilibrium of legal behavior. These arcs of legal change proceed through three intersecting cycles of norm-making and lawmaking. There are iterations among transnational actors in a legal domain as successive IOs attempt to refine or differentiate, compete or complement, normative technologies or products.⁸⁵ There are cycles of lawmaking within nation-states where legislatures, courts, and regulatory agencies put law on their books and then hope for law in action to implement the intended changes—aspirations which frequently lead to implementation gaps, and therefore further rounds of formal lawmaking and implementation until new institutions and behavior settle. Feedback loops exist between the global and the local so that a change in one (e.g., global norms) may lead to change in another (e.g., national statutes), but limitations on implementation lead to another cycle of global norm-making.

These cycles are driven by mechanisms that operate across different areas of the law (e.g., commercial, criminal, civil) and across countries and time. *Actor mismatch* makes settling difficult when actors who are crucial to implementation of law reforms in practice are excluded from the design and bargaining of lawmaking. Excluded actors may choose to nullify reforms until their voices and interests are recognized. *Diagnostic struggles* occur among actors seeking legal change who use competing instruments and draw competing conclusions about the nature of the social, economic, or other problem to be remedied through law. Whose diagnosis comes to prevail frequently determines what prescription gets formalized in the law. *Contradictions* also inhibit settled legal change when lawmakers import competing interests into law without political bargains that properly resolve the disputes. Such contradictions, frequently around competing ideologies and material interests, introduce instability into formal law that easily breaks into open conflict during implementation. Moreover, unresolved contradictions within the law are also likely to produce *indeterminacy* in law’s outcomes, where different actors in practice can draw on competing meanings to justify divergent behaviors.

As the preceding discussion suggests, indicators can be heavily implicated in the recursivity of law. Indeed, the advent, fortunes, and settling (or not) of indicators as normative vehicles are particularly open to a recursive analysis.

Recursive episodes of legal change begin with facilitating conditions that may develop over a long period and then may be precipitated by a crisis, such as the Asian Financial Crisis. IOs either seize the crisis for their own purposes or have it thrust upon them, as did the IMF when the emergent need to combat financing of

⁸⁴ Halliday and Carruthers, “The Recursivity of Law: Global Normmaking and National Lawmaking in the Globalization of Bankruptcy Regimes” (n. 66), 1135–202; Halliday and Carruthers, *Bankrupt: Global Lawmaking and Systemic Financial Crisis* (n. 4).

⁸⁵ Terence C. Halliday, “Recursivity of Global Normmaking: A Sociolegal Agenda,” *Annual Review of Law and Social Science* 6 (2009): 263–90.

terrorism was added to its existing diagnostic instrument on money laundering.⁸⁶ Indicators may be contested or settle or disappear.⁸⁷ It is incumbent upon a theory of indicators to identify when and why episodes that give rise to new indicators begin, and how they conclude, settling at a higher or lower level of acceptance.

Moreover, the institutionalization of indicators, like other forms of legal norms, confronts four mechanisms or processes which influence how and when they settle. In the first place, *diagnostic struggles*, implicitly or explicitly, are at the heart of law reform. The ways that problems are framed substantially determine the steps taken to solve them. As Davis, Kingsbury, and Merry rightly emphasize, indicators are tools that parties can mobilize to spread a patina of science, objectivity, rationality, or systematicity over their claims to define a situation or frame a problem.⁸⁸ Through statistical and graphic representation, their indicator-based claims to offer the “correct” and most refined analysis of the problem are difficult to rebut. Moreover, diagnostic indicators have the operational advantage of translating easily to corresponding remedies. To counter-attack the pseudo-potency of indicators, competing diagnosticians must undermine the validity and reliability of indicators, and the inferences and prescriptions that derive from them. These dynamics generate conflicts among competing epistemologies and their professional bearers. To the extent critics of over-simplified indicators cannot demonstrate their lack of nuance or contingency or salience to the complexities of legal change, then triumphalist bearers of apparently scientific diagnoses will likely contribute to false remedies and failed reforms.

Reliance on indicators also affects patterns of *actor mismatch*. Actors who have the capabilities and resources to gather systematic data can enhance an IO’s claim to a seat at the table of global norm-making or reinforce an interest group’s demand to participate in arenas of national lawmaking. Ordinary lawyers, whose skills and support may be critical to implementation of legal change, will be significantly disadvantaged on this account, in contrast with economists or IFI consultants with their armory of diagnostic instruments and ready funds. Further, since indicators rest on technical expertise, an “expertise barrier”⁸⁹ distorts the representation of those at the indicator-making table who cannot marshal the qualifying competencies. We can expect, therefore, a skewing of reforms towards those actors whose diagnoses include indicators. This skewing may be determinative in the short term, but in the long term will lead to failures of policy implementation, as the non-technocratic players in everyday practice subvert reform initiatives.

The impact of indicators on *contradictions* in national lawmaking is more equivocal. On the one hand, the proliferation of indicators in law reform might

⁸⁶ Michael Levi, “Combating the Financing of Terrorism,” *British Journal of Criminology* 50 (2010): 650–69.

⁸⁷ Davis, Kingsbury, and Merry, “Indicators as a Technology of Global Governance” (n. 2), 87–8.

⁸⁸ *Ibid.*, 83–7.

⁸⁹ Shobita Parthasarathy, “Breaking the Expertise Barrier: Understanding activist strategies in science and technology policy domains,” *Science and Public Policy* 37 (2010): 355–67.

sharpen the edge of contradictory ideologies in lawmaking. Indicators exude a certain precision and instantiate ideology with scientific crispness.⁹⁰ To contradict that ideology, it may be necessary to articulate differences more sharply, even if this articulation does not entail rebutting one set of indicators with another. On the other hand, the potency of indicators may simply overwhelm impressionistic, qualitative, and experiential diagnoses that underlie competing ideologies. Indicators, however questionable their validity or reliability, will wash away opposition and create a false appearance of political settlement.

Indicators may have a perverse effect on the *determinacy* of national lawmaking. Clearly the IFIs intend their indicators to lead to clear prescriptions, institution-building, and substantive law reforms. Thus indicators reinforce voices for the portability of global norms, especially those in the form of model laws that may be slotted into national statute books with a minimum of fuss and disruption. This, again, is likely to produce in practice a false certainty of outcome. The legal transplant literature and anthropology of law reinforces skepticism of “off-the-rack” injections of exogenous law into the complexities and distinctive interpretative communities of legal systems across the world.⁹¹ Applying the same prescriptions in sharply divergent contexts will lead to disruptions, distortions, dislocations, and even disavowal at the local level.⁹²

The recursivity perspective, therefore, leads to a conclusion at sharp variance with the aspirations of IFI legal departments that deploy legal indicators in their global lawmaking enterprise. The distortions to diagnosis, the exclusion of actors who lack the technologies of measurement, the intensification or false resolution of contradictions, and the elusive determinacy of lawmaking at the front end of reform cycles, together increase the likelihood that local resistance, adaptation, and rejection will open up implementation gaps.

In short, the front-end supposed efficiencies apparently provided by indicators will lead to back-end deficiencies. If IFIs insist on imposing their transformational missions on legal systems, they may need better diagnosis and prescription—which in turn require more refined indicators and supplementary information. More certainly, there will need to be further rounds of reform and “recursive change” before any kind of settling occurs that makes sense both to IFIs and their domestic interlocutors.

⁹⁰ Wendy Espeland and Berit Vannebo, “Accountability, Quantification and Law,” *Annual Review of Law and Social Science* 3 (2007): 21–43.

⁹¹ Daniel Berkowitz, Katharina Pistor, and Jean-Francois Richard, “Economic Development, Legality, and the Transplant Effect,” *European Economic Review* 47 (2003): 165–95; Daniel Berkowitz, Katharina Pistor, and Jean-Francois Richard, “The Transplant Effect,” *American Journal of Comparative Law* 51 (2003): 613; Inga Markovits, “Exporting Law Reform—But Will It Travel,” *Cornell International Law Journal* 37 (1995): 95.

⁹² Sally Engle Merry, “Transnational Human Rights and Local Activism: Mapping the Middle,” *American Anthropologist* 108 (2006): 38–51.

Conclusion

Over the past decade, IFI legal departments have expanded their recourse to legal yardsticks in aid of global law reforms. These efforts at legal commensuration and institutional leverage cannot be understood, however, as purely technical instruments for IFIs to monitor, discipline, or reform national and international legal practices. In fact, their use is deeply embedded within the politics of each IFI; they reflect epistemological and status struggles among competing professions, as well as competition among IFIs for survival in an ecology of trade lawmaking and monitoring. Moreover, they function not only as diagnostic instruments, but also as forms of leverage over nation-states. They are part of the armory of national state-building and state reconstruction. Persuasion, in the form of peer group pressure or public shaming, substitutes for the more blatant exercise of conditionality. Commensuration of law disciplines recalcitrant nations and signals disapproval. At best (for IFIs), commensuration drives nation-states into the arms of IFIs for technical assistance and prepares IFI financial decision-makers for sympathetic extensions of loans, grants, and underwriting. At worst, it exacerbates tensions between the global center and the periphery, and intensifies efforts to foil IFI interventions.

If IFI indicators are so inextricably intertwined with the politics of their respective legal departments, the survival of IFIs in global ecologies, and the efficacy of IFIs in national lawmaking, then we should not expect they will be easily expunged. Technical criticism might refine the mechanics of design and execution. But any assault on what has become an integral component of IFI politics will need to answer the perennial rejoinder of indicator designers to their critics: given all their faults, what is the alternative? To social scientists, who are less sanguine about the merits of indicators, it is a fair challenge. The issue is not simply whether there are more valid and reliable ways of appraising legal systems. The issue is whether any alternative to indicators, can meet the ecological, organizational, and professional challenges faced by IFIs and their legal staffs.

This chapter has sought also to extend the rich analysis provided by Davis, Kingsbury, and Merry in *Indicators as a Technology of Global Governance*.

First, to reinforce and extend the argument that indicators are permeated by politics, this chapter proposes that indicators frequently exist within a political ecology of IOs where indicators serve as competitive instruments in struggles for centrality, or even survival, as agents of law reform.

Secondly, to systematize the call for developmental approaches to indicators, this chapter proposes that the rise and fall of indicators, their acceptance or not as measures of law and legal change, can be rigorously approached through the recursive framework for legal change more generally. The recursivity of law provides analytical and theoretical leverage to explain the origins of indicators as reform instruments, the various cycles that attend their subsequent development, and the endings of episodes of indicator-development at one or another equilibrium of behavior.

Thirdly, in both ecological and recursive theory, the legitimacy of IO norm-making is inextricably interwoven with technologies they produce and how they produce them. Sometimes that legitimacy is assured by the appropriation of expert and representative warrants of those participating in the creation and mobilization of indicators. Sometimes it is contained in a rhetorical legitimation that inheres in the name, justification, or application of the indicator.⁹³ At the very least, the emergence of indicators pits one basis of legitimacy—technocratic expertise—against another—representativeness of states or other non-technocratic parties.

Fourthly, the emphasis on expertise, science, and objectivity in the construction of indicators brings the *problematique* of their functions on the global stage into conjunction with the politics of expertise waged by professions to obtain and protect markets⁹⁴ or to assert authority in public policy debates.⁹⁵ Struggles over indicators may reflect a wider ecological positioning of an IO, or an IO's efforts to surmount problems that would forestall its indicator from becoming the global standard. However, they may also be seen as a new terrain of struggle—among professions within IOs and more broadly—to assert what kind of epistemology or expertise will emerge as authoritative in global lawmaking.

⁹³ Halliday, Block-Lieb, and Carruthers, "Rhetorical Legitimation: Global Scripts as Strategic Devices of International Organizations" (n. 5), 1–36.

⁹⁴ Andrew Abbott, "Jurisdictional Conflicts: A New Approach to the Development of Legal Professions" (n. 77), 187–224; Migali S. Larson, *The Rise of Professionalism* (University of California Press, 1977).

⁹⁵ Terence C. Halliday, "The Idiom of Legalism in Bar Politics: Lawyers, McCarthyism, and the Civil Rights Era," *American Bar Foundation Research Journal* (1982): 911–89; Terence C. Halliday, *Beyond Monopoly: Lawyers, State Crises, and Professional Empowerment* (University of Chicago Press: 1987).

9

From Diagnosing Under-immunization to Evaluating Health Care Systems

Immunization Coverage Indicators as a Technology of Global Governance

Angelina Fisher

Introduction

Indicators,¹ which are the focus of several chapters in this book, are rapidly emerging as a new technology of global governance. They can set or become standards, influence the decision-making process of governing entities, and enhance public participation in governance by stimulating contestation of indicators or standards imbedded therein.² At the same time, indicators can conceal nuances and restrict contestation by displacing subjective decision-making with the appearance of objective hard data. Several chapters in this volume discuss the different effects of indicators on behavior of state and non-state actors, both the “governors” and the “governed.”³

This chapter focuses on one set of indicators—immunization coverage indicators produced by the World Health Organization (WHO) and UNICEF. Immunization coverage indicators have been highly successful not only in framing and highlighting the problem of under-immunization but in spurring worldwide efforts to increase the number of children immunized and establish monitoring and reporting regimes worldwide. Moreover, the indicators acquired such prominence that, in the absence of a comprehensive health indicator, one of the immunization coverage indicators—DTP3 (third dose of diphtheria-tetanus-pertussis vaccine)—has come to be used as a proxy for the performance of the overall health system

¹ Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, “Indicators as a Technology of Global Governance,” *Law and Society Review* 36:1 (2012): 73–4.

² *Ibid.*

³ See, e.g., Armin von Bogdandy and Matthias Goldmann, “Taming and Framing Indicators: A Legal Reconstruction of the OECD’s Programme for International Student Assessment (PISA),” this volume, pp. 52–85; Ronen Shamir and Dang Weiss “Semiotics of Indicators: The Case of Corporate Human Rights Responsibility,” this volume, pp. 110–31; René Urueña, “Internally Displaced Population in Colombia: A Case Study on the Domestic Aspects of Indicators as Technologies of Global Governance,” this volume, pp. 249–80.

and subsequently, by international financial institutions (IFIs), as a benchmark that must be fulfilled by a state in order to receive international financing. Despite their broadened use, there has been virtually no public contestation of the immunization coverage indicators.

At first impression, immunization coverage indicators are a simple case of uncontested indicators. Yet, examination of the ecology and dynamics of the indicator production and its transformation from measurement of a narrowly framed problem of under-immunization to vastly broader proxy for the performance of an entire health system reveals a deeper interplay of inter-institutional relations, politics, expertise, and scientific authority that severely restricts the potential for contestation. As the indicator becomes taken for granted, the decisions based on the indicator also receive less scrutiny. In this iterative process, the voices of the people whose lives are affected by the indicators disappear, raising a serious concern that indicator-driven policies will diverge from or impinge upon the needs and rights of the affected population. The chapter draws on the principles of Global Administrative Law (GAL) to address this concern.

The chapter begins by presenting the genealogy of immunization coverage indicators, locating them within the context of WHO's identity-shaping and demand for indicators fueled by the involvement of international development agencies in health reform. The chapter demonstrates that these influences linked immunization coverage indicators inextricably to WHO's institutional identity, which in turn contributed to the appearance of indicators as "highly technical" and "scientific." The chapter proceeds to examine how the immunization indicators are constructed by WHO and used by global actors, and analyzes the indicators as a form of global governance. Particular attention is paid to the concern that limited contestation can result in a chasm between the policies made on the basis of indicators and needs and rights of the people affected by such policies. The chapter concludes by examining a number of mechanisms that can be used to address this concern.

Immunization coverage indicators in context

Immunization is uniformly claimed to be one of the most powerful and cost-effective of all health interventions,⁴ particularly for children, who are more susceptible to and suffer more serious consequences from infectious diseases. The total cost of immunization activities for all 117 low and lower-middle income

⁴ John Maurice and Sheila Davey, "State of the World's Vaccines and Immunization, Third Edition," World Health Organization (WHO), 2009, available at <http://whqlibdoc.who.int/publications/2009/9789241563864_eng.pdf>, xiv; Tracy A. Lieu, Thomas G. McGuire, and Alan R. Hunn, "Overcoming Economic Barriers to the Optimal Use of Vaccines," *Health Affairs*, 24:3 (2005): 666. The impact of widespread immunization is reported to have been so great that the US Centers for Disease Control and Prevention (CDC) put vaccination at the top of its list of 10 great public health achievements of the 20th century and the 2008 Copenhagen Consensus ranked expanded immunization coverage for children fourth on a list of 30 cost-effective ways of advancing global welfare. Panel of distinguished economists convened by the Copenhagen Consensus Center, a Danish think-tank that publicizes the best ways for governments and philanthropists to spend aid and development money, available at <<http://www.copenhagenconsensus.com/Home.aspx>>.

countries for the period 2006–2015 is estimated to be US\$75 billion. The estimated expenditure for immunization activities in 2006–2015 in the 72 poorest countries is US\$35 billion: one-third projected to be spent on vaccines (rising from about US\$350 million in 2005 to nearly US\$1.5 billion per year by 2015) and two-thirds projected to be spent on immunization delivery systems.⁵

In 2002, WHO estimated that 14 percent of the global total mortality in children under five (10 million) was due to diseases that could have been prevented by routine vaccination; in other words, 1.4 million deaths were attributable to vaccine-preventable diseases.⁶ In 2008, the Report by the Secretariat at the Sixty-First World Health Assembly stated that vaccine-preventable diseases were responsible for about 25 percent of the 10 million deaths occurring annually among children under five years of age.⁷ The number of vaccine-preventable deaths among children under five appears to have increased nearly two-fold between 2002 and 2008 (from 1.4 to 2.5 million deaths). The change could have been explained by the expansion of the category of VPDs, introduction of new vaccines, and wider availability of existing vaccines.⁸ However, in the 2009 Global Immunization Coverage report, WHO noted that in 2008, 1.7 million children under the age of five died from vaccine-preventable diseases, bringing the number closer to the

⁵ World Health Organization (WHO), “The cost of immunization programmes in the next 10 years: WHO/UNICEF costing study concludes that with an additional US\$1 billion per year, immunization could save 10 million more lives in a decade,” available at <http://www.who.int/immunization/newsroom/Costing_FINAL.pdf>.

⁶ Of these, 38 percent were attributable to deaths from measles, 27 percent to haemophilus influenzae type B (Hib type B), 20 percent to pertussis, 13 percent to neonatal tetanus, 1 percent to non-neonatal tetanus, and 1 percent to polio, diphtheria, and yellow fever. World Health Organization (WHO) “WHO Vaccine-preventable diseases,” available at <http://www.who.int/immunization_monitoring/diseases/en/>. See also Jennifer Bryce et al., “WHO Estimates of the Causes of Death in Children,” *The Lancet* 365 (2005): 1147–52.

⁷ World Health Organization (WHO), “Global Immunization Strategy: Report by the Secretariat,” April 3, 2008, WHO Doc A61/10, available at <http://apps.who.int/gb/ebwha/pdf_files/A61/A61_10-en.pdf>.

⁸ For example, based on the 2002 pictorial representations, it is clear that in 2000–2003 WHO did not characterize pneumonia and diarrhea as VPDs. This is likely attributable to the fact that vaccines for these diseases were not yet introduced to the market or were not available to everybody. Rotavirus vaccine, which prevents rotavirus—the most common cause of severe diarrhea among infants and young children—was not introduced to the market until 2006. PATH, “Accelerating access to rotavirus vaccines,” available at <<http://www.path.org/projects/rvp.php>>. Prevnar vaccine—which prevents streptococcus pneumonia bacteria—was introduced in 2001 but it was produced from the seven most prevalent strains of the bacteria in the US and thus did not provide protection for children exposed to other strains of the disease. Only in 2009, Synflorix, produced by GlaxoSmithKline, containing 10 serotypes of pneumococcus was introduced to market in Europe. And only in 2010, did Pfizer introduce Prevnar 13, which contains six additional strains (in addition to the seven in the original vaccine), which protect against the majority of the remaining pneumococcal infections. These vaccines are not yet widely accessible by the developing countries. See discussion of Advanced Market Commitments, below. Hib type B, which prevents the type of influenzae that often results in deaths from pneumonia or meningitis, has been available in the US since the 1980s, but it has not been available to the developing countries until 2009 and is still not widely available. It is not clear whether WHO has reclassified pneumonia and diarrhea as VPDs, as more countries are gaining access to Rotavirus and Prevnar.

2002 reported data and inconsistent with the 2008 report.⁹ What this discrepancy illustrates is that numerical simplifications are fragile. Despite the appearance of certainty and objectivity, numerical representations of a complex phenomenon can conceal the construction of categories of measurement on which they depend, in this case the meaning of “vaccine-preventable disease” and the process by which a disease is categorized as such.¹⁰ Whether vaccine-preventable diseases or other diseases bear greater responsibility for childhood mortality shifts focus from one intervention (e.g., vaccination) to another (e.g., provision of clean water). Yet the numbers—107 million infants vaccinated, 2–3 million deaths averted, 1.7 million under-five childhood deaths attributable to vaccine-preventable diseases in one year, 23 million infants not being routinely immunized—conceal the fragility of categorical constructions and instead clearly convey the message that immunization is critical to preventing child mortality and that children’s access to routine vaccination needs to be improved.

Over the years, immunization has attracted the attention of a wide group of actors: governments, international organizations such as WHO, UNICEF, and UNDP, international financial institutions such as the World Bank and Millennium Challenge Corporation, public–private partnerships like GAVI, private foundations like the Bill and Melinda Gates Foundation, and civil society.

In 2000, WHO and UNICEF conducted a retrospective review of all data available on national immunization coverage for a series of antigens from 1980 to 1999. Their intent was to produce indicators that could diagnose the problem of under-immunization at the country level and measure progress toward increasing the number of people who are immunized. Each subsequent year WHO (jointly with UNICEF) continued to produce estimates of immunization coverage for each country. These estimates—or indicators—purport to represent the past performance of a country’s immunization delivery system, are generated through a process that simplifies raw data (i.e., number of children under the age of one who are immunized for a selected set of diseases), and are capable of being used to compare countries, synchronically or over time.¹¹ The annually produced immunization coverage indicators have come to be used for more far-reaching governance purposes, including as a proxy for monitoring the country’s overall health sector performance. Coverage of DTP3 vaccine has become a prominent indicator in the World Bank’s health project financing, The World Bank and International Monetary Fund’s (IMF) poverty reduction initiatives, one of the eligibility indicators for the funding from the Millennium Challenge Corporation (MCC), and even one of the indicators used to assess state effectiveness.

⁹ WHO, “Global Diphtheria-tetanus-pertussis (DTP3) Immunization Coverage, 1980–2009,” July 2010, available at <<http://www.who.int/gho/immunization/en/index.html>>. Incidentally, the 2009 data estimated that the total number of deaths in children under five in 2008 was 8.8 million, making the VPD-attributable deaths accountable for 20 percent of all such deaths.

¹⁰ Sally E. Merry, “Measuring the World: Indicators, Human Rights, and Global Governance,” *Current Anthropology*, 52:3 (2011): S86–7.

¹¹ Davis, Kingsbury, and Merry, “Indicators as a Technology of Global Governance” (n.1), 73–4.

Discussing the role of indicators in global governance, Davis, Kingsbury, and Merry note that

Indicators can be used to evaluate the conduct of states against the standards embodied in the indicator. Those evaluations can in turn form a basis for various actors' decisions on how to create or distribute resources, and how to try to alter the behavior of others or their own behavior. Where a single actor (or set of actors) outside the state have governing power in relation to the state, the governing actors may use indicators in the exercise of this power, for example in taking decisions on whether the state merits particular resources and on whether measures aimed at inducing or achieving compliance with the relevant standard are warranted.

This is precisely what transpires with the immunization coverage indicators. WHO itself acknowledged that

[s]tatistics on levels and trends [of immunization] are used to monitor the performance of services at local, national and international levels. They guide campaigns to eradicate polio, control measles, and eliminate neonatal tetanus. They identify areas of weak system performance and reveal areas where focused attention and extra resources may be required. And, when deciding whether to introduce a vaccine, immunization levels and trends are one important consideration. The numbers for a single vaccine can provide vital information for national and global action. Coverage levels for the diphtheria and tetanus toxoid with pertussis vaccine (DTP), for example, are considered to be a good indicator of health system performance. In addition, some funding agencies, such as the Global Alliance for Vaccines and Immunization (GAVI), often consider immunization coverage levels when reviewing applications for financial and technical support . . .¹²

Immunization coverage indicators thus operate as a technology of global governance.¹³ They influence professional and public opinion towards promotion and expansion of routine immunization, guide allocation of resources and decisions of whether and where to introduce new vaccines, and shape broader health policies. To understand how the indicators became so influential it is important to examine the context within which they arose and gained prominence.

WHO: health statistics expert

In its early years, WHO focused on providing advisory services to countries, mainly through expert advisory panels (by 1957 there had already been 36 such panels) and technical meetings.¹⁴ Initially, its work was directed towards responding to emergencies and working on projects aiming to achieve a specific outcome. Gradually, WHO transitioned to longer-term programming, educational, and training activities, expanded its research activities, and focused on longer-term objectives of

¹² The United Nations Children's Fund (UNICEF), "Immunization Summary: A Statistical Reference Containing Data Through 2009," January 2011, iv–v.

¹³ Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 1), 78–89.

¹⁴ WHO, *The First 10 Years of the World Health Organization*, (Geneva: Palais Des Nations, 1958), available at <http://whqlibdoc.who.int/publications/a38153_%28ch11%29.pdf>, 153.

eradication of communicable diseases.¹⁵ The organization's other major activities were geared to establishing a forum for information exchange—e.g., by creating laboratory networks for reference and exchange of information for influenza and poliomyelitis virus research—and compiling statistical information. Each year, WHO published *Annual Epidemiological and Vital Statistics*.

The International Conference for the Sixth Revision of the International Lists of Diseases and Causes of Death, convened in Paris in April 1948, resulted in the International Statistical Classification of Diseases, Injuries and Causes of Death. The Classification was adopted by the First World Health Assembly in July 1, 1948 together with WHO Regulations No.1, which guided member states in compiling mortality and morbidity statistics by cause, age, and sex for various areas of the national territory and required countries to adopt a form of medical certificate of cause of death.¹⁶ Responsibility for collecting statistics was apportioned between the United Nations, which would deal with collection of population and vital data of general and demographic significance, and WHO, which would deal with those aspects of statistics that involve medical knowledge (certification of causes) and possessing health significance (causes of death).¹⁷ WHO envisioned that “its epidemiological and health statistical services . . . in due course will be one of the main international channels for the transmission of epidemiological and statistical information.”¹⁸

Compilation and computation of statistical data was facilitated by the introduction of computers in the 1960s¹⁹ prompting interest in health statistics on national and local health services.²⁰ WHO began assisting governments in understanding and collecting health data. It also started using statistical techniques in its own research and programs sponsored by WHO. In 1967, 38 WHO-assisted projects on different aspects of vital and health statistics were in operation (as compared to 18 in 1958).²¹ Meetings on health statistics were held frequently, both regionally and internationally and, by the end of WHO's second decade, nearly 230 documents had been issued to national committees to keep them informed of developments in various aspects of health statistics.²² In the 1960s, to reflect the broader range of statistical information now collected, WHO changed the name of its annual publication from *Annual Epidemiological and Vital Statistics* to *World Health Statistics Annual*. Data published in the *World Health Statistics Annual* was

¹⁵ WHO, *The First 10 Years* (n. 14), 170–1.

¹⁶ *Ibid.*, 279.

¹⁷ *Ibid.*, 280.

¹⁸ *Ibid.*, 284.

¹⁹ In 1966, WHO installed its own computer and “more than half the work performed [on it] has been in connexion with health statistics.” WHO, *The Second 10 Years of the World Health Organization* (Geneva: WHO, 1968), available at <<http://whqlibdoc.who.int/publications/14564.pdf>>, 70.

²⁰ Health statistics was a subject of the Nineteenth World Health Assembly in 1966. That year, the Assembly not only adopted a revised *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*, but also requested the Director General to prepare recommendations, definitions, and standards for member states guiding their preparation of health statistics with the goal of improving international comparability.

²¹ WHO, *The Second 10 Years* (n. 19), 71.

²² *Ibid.*, 73.

compiled from information provided to WHO by countries and was issued in three parts: the first volume covered vital statistics and causes of death; the second, data on infectious diseases; and the third, statistics on health personnel and health establishments.²³ As further data was received from countries, it was published in the monthly *Epidemiological and Vital Statistics Report*.²⁴

By the middle of its third decade, WHO “clearly indicated its intention to change health statistics from routine collection and treatment of individual components, in a more or less isolated way, to an integrated approach that met the information needs of health planning and evaluation in all its ramifications.”²⁵ The increased attention given to the subject of health statistics and health planning in turn led to more requests for assistance in this area around the world. A meeting in Geneva in November 1977 on the statistical principles of monitoring and surveillance in public health concluded that WHO could play a vital role in this area by advising on standards, methods, and parameters for assessment, and sponsoring training programmes in monitoring practices.²⁶

Over the years, WHO assembled an impressive health statistics and health information system, publishing a wide range of data and producing guides for various health metrics. Beginning in 1995, WHO started publishing *The World Health Report*, which combined an assessment of global health, including statistics relating to all countries, with a focus on a specific subject. Its goal “is to provide countries, donor agencies, international organizations and others with the information they need to help them make policy and funding decisions.”²⁷

WHO has created the Global Health Observatory (GHO) database, which contains statistics related to Millennium Development Goals (hunger indicators, child mortality indicators, measles immunization coverage, maternal mortality indicators, reproductive health indicators, HIV/AIDS indicators, malaria indicators, TB indicators, water and sanitation indicators, and essential medicines indicators), Mortality and Burden of Disease indicators (life expectancy and global burden of disease indicators), World Health Statistics (data on mortality and burden of disease, cause-specific mortality and morbidity, selected infectious diseases, health service coverage, risk factors, health expenditure, health inequities, and demographic and socioeconomic statistics), and statistics on Violence and Injuries.²⁸ According to WHO’s Programme on Health Statistics, WHO-produced figures “carry great weight in national and international resource allocation, policy making and programming, based on its reputation as ‘unbiased’ (impartial and

²³ The third volume, published in 1968, included for the first time charts showing the percentage distribution of populations, health personnel, and hospital beds in urban and rural areas. WHO, *The Second 10 Years* (n. 19), 73–5.

²⁴ *Ibid.*, 73.

²⁵ WHO, *The Third 10 Years of the World Health Organization*, (Geneva: WHO, 2008), available at <http://www.who.int/global_health_histories/who-3rd10years.pdf>, 275.

²⁶ WHO, *The Third 10 Years* (n. 25), 281.

²⁷ WHO, “The World Health Report: Introduction,” available at <<http://www.who.int/whr/en/index.html>>.

²⁸ WHO, “Global Health Observatory Data Repository,” available at <<http://apps.who.int/ghodata/>>.

fair), global (not belonging to any camp), and technically competent (consulting leading research and policy institutions and individuals)."²⁹

Most recently, WHO overhauled its website, populating it with statistics-heavy reports and reorganizing the information to make it significantly more user-friendly. *WHO Vaccine-Preventable Diseases: Monitoring System 2010 Global Summary* provides Immunization Profiles for each country, in which WHO collates information on demographics, number of reported cases of eleven diseases (vaccination against which is included in the routine vaccination package), percent of target population vaccinated (by antigen, disaggregated by country-reported data, survey-based data, and WHO estimates), the country's immunization schedule, an extensive list of immunization indicators, which includes data relating to planning and management, system performance (coverage indicators by antigen using number of districts as the unit of comparison), disease surveillance, vaccine supply, vaccine wastage (national percentage wastage for each antigen), vaccine safety, finance, introduction of new and under-utilized vaccines, maternal and neonatal tetanus elimination, protection at birth against neonatal tetanus, polio eradication, vitamin A deficiency status, and yellow fever risk and vaccine inclusion in the vaccine schedule. WHO also provides regional and global immunization profiles along these variables and presents country-reported data on immunization coverage by antigen as well as "selected indicators of immunization system performance" in a manner that allows for comparisons both of a country's performance over time as well as performance as among countries.³⁰

WHO's focus on developing an expertise in health statistics was fueled by the demand for "measurable" outcomes and easily comparable information. At the same time, it became part of WHO's business model that helped shape the organization's identity in the face of the growing number of actors in the global public health arena.

Demand for indicators: from immunization to health

In the early 1950s, WHO was approached by a group, convened by the United Nations with the participation of some of the specialized agencies, concerned with

²⁹ WHO, "WHO Programme on Health Statistics," available at <<http://www.who.int/healthinfo/statistics/programme/en/index.html>>.

³⁰ Selected indicators are: (i) does the country have a multi-year plan (MYP) for immunization?; (ii) % of districts ≥ 90 percent MCV1 coverage; (iii) % of districts > 10 percent DTP3- DTP1 drop-out rates; (iv) national % wastage for Hib; (v) national % wastage for MCV % coverage reports received at national level from districts compared to number of reports expected; (vi) % of immunization spending financed using Government funds; (vii) number of districts with microplans that include activities to raise immunization coverage; (viii) % of districts with microplans including activities to raise immunization coverage; (ix) % of districts ≥ 80 percent DTP3 coverage; (x) was there any district with vaccine supply interruption?; (xi) % of districts supplied with sufficient number of AD syringes; WHO, "WHO Vaccine-Preventable Diseases: Monitoring System—2010 Global Summary" WHO Doc WHO/IVB/2010, available at <http://www.who.int/immunization/documents/who_ivb_2010/en/index.html>; WHO, "Indicators of Immunization System Performances," WHO Doc WHO/V&B/WHO/IVB/2010, available at <http://whqlibdoc.who.int/hq/2010/WHO_IVB_2010_eng_pR243-end.pdf>.

the description of social development or defect as a guide to the establishment of international programs of assistance. The group wanted "some means of comparing different levels or standards of living that would make it possible to form a better judgement as to what action was required to raise them and to assess how far such action had been successful."³¹ A United Nations committee of experts decided that it was impossible to prescribe a single index by which a level of living could be defined, but suggested looking to components that were "clearly defined" and "could be measured": health, food, education, conditions of work, and employment.³² The means of measuring these components were called "indicators" and WHO was asked to advise on indicators for health. The study group set up by WHO ultimately could not recommend a direct measurement of social well-being, but in 1956 the WHO Expert Committee on Health Statistics endorsed the opinion that the expectation of life at birth, at one year, or at any other age, was theoretically the best health indicator. However, since it was available for only a small number of countries the committee recommended that "a proportional mortality ratio (percentage of total deaths represented by deaths at the age of fifty and over) be experimentally taken as a comprehensive health indicator."³³ The need for a comprehensive health indicator likely subsided during the 1960s and 1970s, when the international community changed its approach to global health from promoting primary health care and "Health for All in the Year 2000" (as defined during the 1978 conference in Alma-Ata)³⁴ to promoting *Selective* Primary Care, which advocated for massive coverage of a few key interventions that address the most important causes of child mortality and morbidity, identified by the international community initially as growth monitoring, oral rehydration, breast-feeding, immunization, and then also food supplementation, family planning, and female education.³⁵

Communicable (or infectious) diseases already have been a constant item on WHO's agenda,³⁶ but during the 1970s WHO's attention to immunization heightened. Building on the success of the smallpox eradication campaign,³⁷ in

³¹ WHO, *The First 10 Years* (n. 14), 284.

³² *Ibid.*, 284.

³³ *Ibid.*, 285.

³⁴ Theodore M. Brown, "The World Health Organization and the Transition from 'International' to 'Global' Public Health," *American Journal of Public Health* 96 (2006): 67.

³⁵ Some of the theories regarding the shift in focus are laid out in the Report of the UN Millennium Project on the Task Force on Child Health and Maternal Health, 2005, available at <<http://www.unmillenniumproject.org/documents/maternalchild-complete.pdf>>, 38.

³⁶ In 1955, WHO initiated a campaign of malaria eradication. It was conceived and promoted with much enthusiasm under the belief that malaria eradication "would usher in economic growth and create overseas markets for US technology and manufactured goods," "build support for local governments and their US supporters," and "help win 'hearts and minds' in the battle against Communism." In 1959, at the urging of the former Soviet Union, the World Health Assembly began a program of global smallpox eradication. Ultimately, the malaria campaign failed and in 1969, the World Health Assembly declared that eradication of malaria was not feasible in many parts of the world and emphasized the need to develop rural health systems and integrate malaria control into general health services. Brown, "The World Health Organization" (n. 34), 65.

³⁷ In 1967, WHO launched the Intensified Smallpox Eradication Program and, in 1979, WHO certified eradication of smallpox.

1973, WHO initiated the Expanded Programme on Immunization (EPI). Immunization had already been identified as a priority in maternal and child health activities in the early 1960s and immunization records were available for at least some of the countries, but with the adoption of Resolution WHA 27.57 in 1974, all states were now called upon to “develop or maintain immunization and surveillance programmes against some or all of the following diseases: diphtheria, pertussis, tetanus, measles, poliomyelitis, tuberculosis, smallpox and others, where applicable.” The program and budget statement for 1975 mentioned that “[i]mmunization is the most effective and rapidly applicable measure of preventive medicine available to health authorities”³⁸ and the Joint UNICEF/WHO Committee on Health Policy agreed that both WHO and UNICEF should renew their interest and double their efforts in assisting countries in extending immunization to their children, thus further ensuring that immunization will feature on countries’ agenda. From the late 1970s, countries began reporting data on national immunization coverage, first separately to the World Health Organization (WHO) and UNICEF, and since 1998 on a WHO/UNICEF Joint Reporting Form on Immunization (JRF).

The 1980s saw WHO go through turmoil. The World Bank’s influence in the global health arena grew and the Bank began to make direct loans for health services.³⁹ At the same time, the World Health Assembly froze WHO’s budget in 1982 and, in 1985, the United States withheld its contributions to WHO.⁴⁰ WHO began relying on donations by multilateral agencies and “donor” nations⁴¹ and by the beginning of the 1990s, WHO’s extrabudgetary funding surpassed the regular budget of the organization by \$21 million, contributing 54 percent of WHO’s overall budget.⁴² This meant that while WHO’s priorities were formally set by the World Health Assembly, wealthy donors had a lot of influence.⁴³ While WHO was juggling to balance its programs with the desires of the donors, the World Bank stepped prominently into the health care arena. By 1996, the World Bank’s lending portfolio in health, nutrition, and population had reached \$13.5 billion, far greater than WHO’s budget.⁴⁴

In 1998, newly elected Director General Gro Harlem Brundtland, determined to restore the role of WHO in the global arena, established a Commission on Macroeconomics and Health, chaired by Jeffrey Sachs and composed of former

³⁸ WHO, *The Third Ten Years* (n. 25), 191.

³⁹ Brown, “The World Health Organization” (n. 34), 67.

⁴⁰ The withdrawal of funds was in protest against WHO’s “Essential Drug Program,” which was opposed by US pharmaceutical companies. Fiona Godlee, “WHO in Retreat: Is It Losing Its Influence?,” *British Medical Journal* 309 (1994): 1492.

⁴¹ Gill Walt, “WHO Under Stress, Implications for Health Policy,” *Health Policy* 24 (1993): 125–44.

⁴² Brown, “The World Health Organization” (n. 34), 68.

⁴³ A Danish study assessing the work of 11 agencies, including WHO, at country level in Kenya, Nepal, Sudan, and Thailand concluded that political considerations determined the allocation of WHO’s regular program funds, both between countries and between programs. DANIDA, *Effectiveness of Multilateral Agencies at Country Level* (Copenhagen: CowiConsult-Danida, 1991).

⁴⁴ Brown, “The World Health Organization” (n. 34), 68.

ministers of finance, representatives from the World Bank, the IMF, the World Trade Organization, the UNDP, and public health leaders. In its report, published in 2001, the Commission recommended scaling-up access of the world's poor to essential health services by focusing on specific interventions. Control of communicable diseases and improved maternal and child health were identified as the highest public health priorities⁴⁵ and the Commission placed a heavy emphasis on the connection between health generally, and eradication of communicable diseases in particular, to economic benefits and poverty reduction.⁴⁶

Beginning in the 1990s and continuing through the first decade of the new millennium, international financial institutions (IFIs) came under heavy criticism for their use of conditionality in development funding. At the same time, donors and citizens calling for better aid stepped up demands for scrutiny of development funds.⁴⁷ These developments caused IFIs to shift their lending practices from ex ante to performance- or outcome-based conditionality. In 2002, IMF's "Guidelines on Conditionality" noted that "program-related conditions may contemplate the member meeting particular targets or objectives (outcomes-based conditionality)."⁴⁸ The World Bank also highlighted the need to "sharpen the focus on achieving results in supporting developing countries' poverty reduction strategies and contribution toward the Millennium Development Goals."⁴⁹ The European Commission also reformed its use of conditionality, conditioning its disbursement of aid not on the implementation of specific policy measures but on outcomes in key economic and social sectors.⁵⁰ These changes meant that development agencies were now looking for "performance" indicators. Since improving health systems was seen as necessary for poverty reduction, demand for health indicators intensified. WHO's established expertise in statistics meant that the organization was well positioned to reclaim its role in global public health.

In 2000, the World Health Organization published the *World Health Report 2000—Health Systems: Improving Performance*, wherein it measured and ranked health systems in 191 member states on the basis of five indicators: overall level of population health; health inequalities (or disparities) within the population; overall level of health system responsiveness (a combination of patient satisfaction and how well the system acts); distribution of responsiveness within the population (how well people of varying economic status find that they are served by the health

⁴⁵ Report of the Commission on Macroeconomics and Health: Investing in Health for Economic Development, December 20, 2001, available at <<http://whqlibdoc.who.int/publications/2001/924154550x.pdf>>, 20>.

⁴⁶ Howard Waitzkin, "Report of the WHO Commission on Macroeconomics and Health: A Summary and Critique," *The Lancet*, 361 (February 8, 2003): 523–6.

⁴⁷ Rome Declaration on Harmonization (February 2003), Paris declaration on Aid Effectiveness: Ownership, Harmonisation, Alignment, Results and Mutual Accountability (March 2005); "Outcome-based Conditionality: Too Good to be True?" *Eurodad Report* (February 2008); Christopher Adam and Jan Willem Gunning, "Performance-Based Conditionality: A European Perspective," *World Development* 32 (2004): 1059–70.

⁴⁸ Timothy Geithner, "The Modalities of Conditionality," International Monetary Fund (IMF), January 8, 2002.

⁴⁹ World Bank, "From Adjustment Lending to Development Policy Lending," World Bank, 2004.

⁵⁰ Adam and Gunning, "Performance-Based Conditionality" (n. 47), 1059–70.

system); and the distribution of the health system's financial burden within the population (who pays the costs).⁵¹ Rather than being welcomed as much needed indicators of health system, the rankings were met with outcry and heated debates.⁵² Although some critics challenged the reports' methodology, much of the opposition to the rankings was clearly politically driven (the fact that the United States scored 37th out of 191, behind countries like Saudi Arabia, Singapore, Andorra and Dominica, fueled the criticism of the report).⁵³ In response, WHO Director General established a technical consultation process and a series of expert meetings in all WHO regions.⁵⁴ In 2002, a report by the Secretariat to WHO's Executive Board recommended that in subsequent reports information be presented "alphabetically with countries assigned into several groups (eg from A to E)."⁵⁵ WHO continues to produce World Health Reports and provide health statistics data for all member states, but it no longer publishes an overall score for countries' health system performance and does not rank countries in any category.

As early as 1998, WHO proposed that coverage for DTP3 antigen be used as one of the criteria for allocating funds to countries⁵⁶ and in 2000, a discussion paper issued by the Health, Nutrition and Population Family (HNP) of the World Bank's Human Development Network, argued that DTP3 coverage data was a good "indicator of a health system's capacity to deliver essential services to the most vulnerable members of a population."⁵⁷ Since production of a comprehensive

⁵¹ WHO, World Health Organization Assesses the World's Health Systems, available at <http://www.who.int/whr/2000/media_centre/press_release/en/>.

⁵² See, e.g., Vincent Navarro, "Assessment of the World Health Report 2000," *The Lancet* 356 (2000): 1598–601; Celia Almeida et al., "Methodological Concerns and Recommendations on Policy Consequences of the World Health Report 2000," *The Lancet* 357 (2001): 1692–7; Kjeld Miller Pedersen, "The World Health Report 2000: Dialogue of the Deaf," *Health Economics* 11 (2002): 93–101; Eeva Ollila and Meri Koivusalo, "The World Health Report 2000: World Health Organization Health Policy Steering Off Course—Changed Values, Poor Evidence, and Lack of Accountability," *International Journal of Health Services* 32 (2002): 503–14.

⁵³ Raisa Deber, "Why Did the World Health Organization Rate Canada's Health System as 30th? Some Thoughts on League Tables," *Healthcare Quarterly* 7 (2004); Glen Whitman, "WHO's Fooling Who? The World Health Organization's Problematic Ranking of Health Care Systems," *CATO Institute Briefing Papers*, February 28, 2008; Robert B. Helms, "Sick List: Health Care a la Karl Marx," *Wall Street Journal* (Europe), June 29, 2000, 8.

⁵⁴ WHO, "Health Systems Performance," available at <<http://www.who.int/health-systems-performance>>.

⁵⁵ Report of the Secretariat on the Assessment of Health Systems' Performance, December 12, 2002, UN Doc EB11/6.

⁵⁶ WHO, "Report of the Meeting of the Scientific Advisory Group of Experts (SAGE)," WHO Doc WHO/GPV/98.06, June 9–11, 1998, available at <http://whqlibdoc.who.int/hq/1998/WHO_GPV_98.06.pdf>.

⁵⁷ Eduard Bos and Amie Batson, "Using Immunization Coverage Rates for Monitoring Health Sector Performance: Measurement and Interpretation Issues," The World Bank, August 2000, available at <<http://siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/Resources/281627-1095698140167/Bos-UsingImmunization-whole.pdf>>. The authors argued that: (i) immunization is a health output with a strong impact on child morbidity, child mortality, and permanent disability; (ii) the target group consists of zero-to-one-year-old children, and members of the group consist of the cohort of children born each year. Therefore, immunization coverage can provide timely evidence of improvement and deterioration (sensitive indicator); (iii) measurement of immunization coverage can be relatively straightforward and inexpensive, and results in valid and verifiable information; (iv) definitions used in surveys and health information systems to measure

health system indicator proved to be politically impossible, DTP3 coverage—coverage of the third dose of DTP vaccine—became the best available proxy. Importantly, effective coverage does not measure the impact of the intervention; it merely represents “an intermediate step in achieving a health impact.”⁵⁸ For example, high coverage of DTP vaccine does not indicate that a large proportion of the population is in fact protected against diphtheria, pertussis, and tetanus, but it does indicate that a high number of targeted individuals were given the vaccine. The impact, which is a function of multiple factors, including the timing of vaccination and proper vaccine storage, could be measured more accurately by conducting a biological assessment of whether antibodies are present in each person. Arguably, measuring the impact of vaccination is a more accurate assessment both of the intervention itself and of a health system at large. A country’s high coverage of individuals with antibodies against a particular disease in their blood system would evidence availability and access to health facilities and providers equipped with necessary skills and technology, contact of population with health providers (presumably blood would be drawn during an annual physical exam or a sick visit),⁵⁹ availability of effective vaccines, proper cold storage facilities, access to vaccines, and scientific knowledge required to analyze the blood serum. This certainly seems a more comprehensive proxy for the health care system than simply looking at the coverage of immunization. Yet, measuring the impact of vaccination in this manner is very costly and probably unwarranted except where there are real possibilities of vaccine ineffectiveness. Furthermore, it requires facilities and personnel that may not be available in many countries.

DTP3 coverage indicator was considered a good proxy because it signaled (i) availability of health providers to administer vaccines; (ii) existence of and accessibility to a health facility; (iii) population knowledge of immunization; and (iv) availability of follow-up care (because the child had to be brought to the clinic on three separate occasions). This indicator was also attractive because it is relatively easy to measure, it is viewed as not dependent on exogenous factors, it is not subject to targeted campaigns, it is produced annually, and it is widely available. The data for DTP3 was widely available because WHO’s historic emphasis on routine immunization coupled with its established identity as a health statistics expert and a clearing house for health data ensured that (a) countries had good infrastructure to monitor and report on routine immunization coverage, which includes

immunization coverage can be precise and objective so as to enable comparisons across countries and over time; (v) immunization is a universally recommended and cost-effective public health priority, for which international adopted targets exist; (vi) immunization coverage rates are frequently available at the sub-national level and thus can be used to monitor impact and for targeting of service delivery; and (vii) immunization coverage rates are useful to monitor progress in expanding essential health services in adverse health settings, and as “safeguard” indicators when health system reforms are changing delivery or financing of health services in settings in which immunization coverage is already high.

⁵⁸ WHO, “Background Paper for the Technical Consultation on Effective Coverage of Health Systems,” August 27–29, 2001, 14.

⁵⁹ Although countries could, in theory, mandate people to appear for a blood draw, there is no benefit to the individual that would incentivize them to comply. Unless, of course, threat or coercion was used.

DTP3 vaccine; (b) countries got into the habit of routinely reporting immunization coverage; and (c) WHO (in partnership with UNICEF) developed the capabilities for collecting, collating, and converting the reported data into an indicator through “scientific” methodology, giving the indicator the appearance of being “objective,” “scientific,” and “reliable.”

Over the years, the DTP3 coverage indicator became widely used—both as an indicator of immunization system performance and a proxy for health system performance—by policymakers, advocates, governments, international financial institutions, and donors for purposes of resource allocation, determining whether or not to introduce a new vaccine into the country, health system funding, and advocacy focus. The next part of the chapter examines the production of routine immunization coverage (among which DTP3 is one of the monitored antigens) and the use of DTP3 coverage.

Immunization coverage indicators: Production and use

Production of the indicators

Coverage of an intervention—for example, immunization—is an indicator of service provision function. It represents the ratio of number of population units who have received an effective intervention to the number of population units in need of intervention.⁶⁰ Data on immunization coverage can be collected through reporting by vaccine providers and aggregated at different levels (e.g., village, district, region, state, etc.) or it can be obtained through random sampling or surveys.

WHO and UNICEF generate immunization coverage indicators using administrative data (i.e., data that is reported from health facilities up the chain) as well as data from surveys.⁶¹ Administrative data are usually based on a country’s routine monitoring system. Service providers report on the number of vaccinations administered⁶² during a given period to local public health authorities.⁶³ The data are then aggregated and reported to the next administrative level and subsequently aggregated, analyzed, and used at national level. National coverage data are reported annually to WHO and UNICEF on the JRF.⁶⁴ Sometimes, countries send their data to regional health organizations, which subsequently forward them to WHO/UNICEF.

⁶⁰ Christopher J.L. Murray and David B. Evans, *Health Systems Performance Assessment: Goals, Framework and Overview* (Geneva: WHO, 2003), 9; Christopher J.L. Murray and David B. Evans, *Health Systems Performance Assessment: Debates, Methods and Empiricism* (Geneva: WHO, 2003), 11.

⁶¹ Surveys are frequently used in conjunction with administrative data, but may also constitute the sole source of information. WHO, “WHO Vaccine-Preventable Diseases: Monitoring System—2010 Global Summary,” WHO, 2010.

⁶² A vast majority of countries maintain a system whereby a health provider records vaccinations as they take place and tallies them up. Interview with Interviewee 1, July 7, 2010. On file with author.

⁶³ Anthony Burton et al., “WHO and UNICEF Estimates of National Infant Immunization Coverage: Methods and Processes,” *Bulletin of the World Health Organization* 87 (2009): 536.

⁶⁴ Burton, “WHO and UNICEF Estimates” (n. 63), 536.

The Joint Reporting Process can be summarized as follows:

- (i) WHO and UNICEF Headquarters prepare the JRF. Regional organizations are consulted and specific questions may be included in the JRF for specific countries.⁶⁵
- (ii) Once the JRF is agreed on, it is communicated to UNICEF national offices and WHO national offices (via WHO regional offices) and subsequently to the country's ministry of health.
- (iii) Data is collected, aggregated, and analyzed at the national level and reported back to WHO and UNICEF, often through the organizations' regional and national offices. Additional data may be solicited from countries in the course of preparation of WHO annual reports.

Review of the country's reported data takes place in a form of either data quality self-assessment (DQS) or data quality audit (DQA). DQS is intended to be used by countries to evaluate different aspects of the immunization monitoring system at district and health unit levels. Specifically, DQS aims to determine the accuracy of reported numbers of immunizations, and the quality of the immunization monitoring system. The purpose is to assist countries in diagnosing problems and improve district monitoring. WHO does not publish information related to the accuracy of a country's data.

Once the information from the countries is received by WHO, it is submitted to the process of producing estimates of immunization coverage, presented as a percentage of target population that has been vaccinated. What is known about the methodology is summarized below:

- (i) each country's data are reviewed individually;
- (ii) no correlation of immunization coverage is made with income level, development status, population size, or geography;
- (iii) if survey data (e.g., Expanded Programme on Immunization cluster survey,⁶⁶ UNICEF Multiple Indicators Cluster Survey,⁶⁷ and the Demographic and Health Survey)⁶⁸ confirm within ± 10 percentage points reported data, reported data is used;

⁶⁵ For example, the Pan-American Health Organization was interested in evaluating how immunization was administered by ethnic groups so it included a question to that effect in JRF for some countries in the region. Interview with Interviewee 2, July 19, 2010.

⁶⁶ WHO, "Immunization Coverage Cluster Survey, Reference Manual," June 2005, WHO Doc WHO/IVB/04.23.

⁶⁷ UNICEF, "Multiple Indicator Cluster Surveys/MICS3," May 2009, available at http://www.childinfo.org/mics3_background.html

⁶⁸ Demographic and Health Surveys (DHS), "Measure DHS, Demographic and Health Surveys," DHS, 2008, available at <http://www.measuredhs.com/aboutsurveys/dhs/start.cfm>. DHS model questionnaires are available at http://www.measuredhs.com/pubs/search/search_results.cfm?Type=35&srchTp=type&newSrch=1 without registration. Countries' final DHS reports are available free of cost, also without registration, at http://www.measuredhs.com/pubs/search/search_results.cfm?Type=5&srchTp=type&newSrch=1.

- (iv) if multiple survey points show “a fairly consistent” relationship with the reported data trend *and* survey data are “significantly” different from reported data, then reported data “calibrated to the level established by the survey data” is used;
- (v) if survey data is “inconsistent” with reported data *and* there is “no consistent relationship” between reported data and survey data *and* survey data “appear more reliable,” then survey data is used (“with interpolation between survey data points for intervening years”);
- (vi) if multiple data points are available for country, vaccine/dose, year, then “potential biases in each source are considered and an attempt is made to construct a consistent pattern over time from the data with the least potential for bias consistent with temporal trends and comparisons between vaccines” (i.e., no averages are used);
- (vii) if coverage patterns are inconsistent with the vaccine and dose numbers, “an attempt to identify and adjust for possible biases is made”;
- (viii) when several estimates are possible “alternative explanations that appear to cover the observed data are constructed and treated as competing hypotheses”;⁶⁹
- (ix) where data are missing prior to the first-ever reported coverage, the coverage is assumed to be 0 percent; where data are missing between two years where coverage was reported, the coverage estimate is a “linear interpolation of the two reported coverage rates”; if the coverage has ceased to be reported, it is assumed that coverage in the years following the last report will remain at the same level as was last reported.⁷⁰

Coverage estimates are usually calculated for each vaccine and for the number of doses received. The target population depends on type of vaccine, the dose, and national policies. Immunization estimates reflect vaccinations given during routine immunization services, regardless of whether the recommended minimum age and the minimum interval between multi-dose vaccines is respected.⁷¹ They do not, however, include vaccinations received during supplemental vaccination campaigns.⁷²

Once computed, draft estimates are sent to each national authority for feedback. Comments received are reviewed by WHO and UNICEF working groups and, if necessary, estimates are modified accordingly.⁷³

Uses of the indicators

The DTP3 coverage indicator (published by WHO for a particular country or region) has been used extensively, although predominantly at the global level.

⁶⁹ Burton, “WHO and UNICEF Estimates” (n. 63), 537–8.

⁷⁰ WHO, “WHO Vaccine-preventable Diseases” (n. 6).

⁷¹ Burton, “WHO and UNICEF Estimates” (n. 63), 537; “WHO, Vaccine-preventable Diseases” (n. 6).

⁷² Ibid.

⁷³ Ibid.

International organizations use it to evaluate their own work, to motivate donors by demonstrating the impact of their investment in supporting immunization, and to plan further activities.⁷⁴ For example, the indicator features prominently in the WHO/UNICEF Global Immunization Vision and Strategy (GIVS).⁷⁵ PATH, an international nonprofit organization, has used WHO data on state-level immunization coverage in India to evaluate the impact of its interventions.⁷⁶ Even at country level, national agencies may consider WHO/UNICEF estimates as a “check” on their system: a national EPI manager at the Ghana Health Service noted that WHO immunization estimates are helpful because “independent opinion is important in strengthening systems,” otherwise “it’s like playing your own match and referring yourself.”⁷⁷

Donor organizations that focus specifically on delivery of immunization may use the immunization coverage indicator as a sole indicator while organizations that focus more broadly on improving health systems or reducing poverty may use immunization coverage indicators as part of a cluster. For example, GAVI uses the DTP3 coverage indicator as an eligibility indicator for access to funding for new vaccines. Countries with less than 70 percent⁷⁸ DTP3 coverage are *not* eligible for GAVI funding supporting introduction of some of the new vaccines such as Haemophilus influenza type b, pneumococcal vaccine, and rotavirus vaccine (although they remain eligible for funds supporting the introduction of vaccines against Japanese encephalitis (JE), Meningitis A, and yellow fever, for cash-based support programs, and for funds supporting (i) improving immunization services; (ii) strengthening of health systems; (iii) strengthening of civil society organizations; and (iv) improving injections).⁷⁹ The DTP3 coverage indicator is also used as a mark of eligibility for participation in the Advanced Market Commitment (AMC) project founded in 2009 by a consortium of governments (Italy, United Kingdom, Canada, Russia, and Norway), an international donor organization (Bill & Melinda Gates Foundation), and GAVI Alli-

⁷⁴ See, e.g., Philippe Duclos, Jean-Marie Okwo-Bele, and Marta Gacic-Dobo, “Global Immunization: Status, Progress, Challenges and Future,” *BMC International Health and Human Rights* 9 (October 2009): Suppl 1.

⁷⁵ “By 2010 or earlier (i) *Increase coverage*. Countries will reach at least 90 percent national vaccination coverage and at least 80 percent vaccination coverage in every district or equivalent administrative unit. (ii) *Reduce measles mortality*. Globally, mortality due to measles will have been reduced by 90 percent compared to the 2000 level. . . . By 2015 or earlier (as the case may be) (i) *Sustain coverage*. The vaccination coverage goal reached in 2010 will have been sustained. (ii) *Reduce morbidity and mortality*. Global childhood morbidity and mortality due to vaccine-preventable diseases will have been reduced by at least two thirds compared to 2000 levels. . . .” WHO, “Global Immunization Vision and Strategy,” January 8, 2009, available at <<http://www.who.int/immunization/givs/goals/en/index.html>>.

⁷⁶ Interview with Interviewee 3.

⁷⁷ Interview with Interviewee 4.

⁷⁸ GAVI Alliance, “Guidelines on Country Proposals—For Support for: New and Underused Vaccines,” GAVI Alliance, 2011, available at <<http://www.gavialliance.org/support/apply>>.

⁷⁹ Comprehensive information on GAVI’s support for new and underused vaccines is available at <<http://www.gavialliance.org/support/apply>>.

ance partners (World Bank, UNICEF, and WHO) to stimulate the creation of markets for vaccine against pneumococcal disease.⁸⁰

Countries apply to GAVI for AMC funding according to GAVI standard procedures, which means they must meet the 70 percent DTP3 coverage to qualify. The reason for the threshold appears to be driven by donors who feel that since low DTP3 coverage indicates that countries are not able to handle routine immunizations, it would be inefficient and even wasteful to supply new vaccines to such countries. As the Senior Public Health Specialist at the Health, Nutrition and Population Department of the World Bank explained, “new vaccines come at a high cost not only in terms of vaccine price but also in terms of what is required of the immunization system. For example, countries may need to maintain good cold chain facilities, have storage capacity, etc. So when we introduce a new vaccine to a country we want to make sure that it does not further strain the system.”⁸¹ Another interviewee, in the advocacy sector, put it more bluntly, “donors, like USAID do not want to parachute expensive products into inefficient systems.”⁸²

In the world of development financing, the use of indicators generally has been welcomed as part of the move to *ex ante* vs *ex post* conditionality. While the discussion of the merits of *ex ante* vs *ex post* conditionality is beyond the scope of this chapter,⁸³ the change has been an important driver for the use of indicators in development financing and has spurred the need for credible indicators. In health reform financing, a Senior Public Health Specialist at the Health, Nutrition and Population Department of the World Bank, noted that one of the reasons the World Bank uses immunization coverage indicators rather than data reported by the states is because the information is “vetted,” the presumption being that whatever biases are present in country data are removed in the WHO’s estimation process.⁸⁴ The Millennium Challenge Corporation uses WHO’s indicators because they fit the Corporation’s criteria of being developed by a third party that uses “an analytically rigorous methodology and objective and high-quality data.”⁸⁵

The World Bank uses the DTP3 coverage indicator as one of the performance indicators in International Development Association (IDA) funding⁸⁶ and in a

⁸⁰ Pursuant to AMC, donors committed \$1.5 billion to guarantee the price of the vaccines once it is developed. GAVI endorsed an additional budget of up to \$1.3 billion for the period 2010–2015 to help fund the cost of vaccines. The initiative is aimed at creating the potential for a viable market and the monetary commitment incentivizes vaccine makers to invest in research, train staff, and build manufacturing facilities. In exchange, participating companies sign legally binding agreements to supply their vaccines at a price no higher than \$3.50 for 10 years. The \$3.50 price represents a 95 percent reduction from the current pricing in industrialized countries. GAVI and GAVI eligible countries will pay the long-term price of the vaccines. For information on Pneumococcal AMC, see information and documents at <<http://www.gavialliance.org/funding/pneumococcal-amc/>>.

⁸¹ Interview with Interviewee 5.

⁸² Interview with Interviewee 6.

⁸³ Celine Tan, *Governance through Development* (Taylor & Francis, 2011).

⁸⁴ Interview with Interviewee 5.

⁸⁵ <<http://www.mcc.gov/mcc/selection/indicators/index>>.

⁸⁶ International Development Association (IDA), “Bolivia: Better Access to Essential Health Services,” September 2008, available at <<http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21940429-pagePK:51236175-piPK:437394-theSitePK:73154,00.html>>.

cluster of indicators as key outcomes and/or as intermediate measures of progress in lending for improvements of health systems or the reduction of poverty.⁸⁷ For example, Heavily Indebted Poor Countries (HIPC), a group of 40 least developed countries with high levels of poverty and debt overhang, are eligible for special assistance from the IMF and the World Bank. To be considered for HIPC Initiative assistance, in addition to meeting financial eligibility requirements, a country must develop a Poverty Reduction Strategy Paper (PRSP).⁸⁸ Adoption and implementation of PRSP for at least one year is one of the conditions a country must fulfill in order to receive full and irrevocable reduction in debt available under the HIPC Initiative (completion stage).⁸⁹ As of 2005, 19 of the 20 final PRSPs for countries in Africa included DTP3 immunization coverage as an indicator of progress toward poverty reduction in the social sector.⁹⁰ The immunization DTP3 coverage indicator is also used as part of a cluster of qualifying indicators by the Millennium Challenge Corporation.

Such use of the immunization coverage indicators has real impacts on national policies. A Project Appraisal Document on a proposed Adaptable Program Credit from the World Bank to Bolivia (in support of the second phase of the health sector reform program) noted the following as evidence of success: "... The baseline [defined as the number of municipalities with DTP3 coverage of less than 80 percent] was very high, with 67 percent of municipalities showing low coverage of immunization. *During 1999 and 2000, a huge effort was made to expand immunization to remote locations and by end-2000, the target for 2001 had been greatly exceeded, with only 21 percent of municipalities showing low coverage*" (emphasis added).⁹¹ The reason for this success is not surprising: DTP3 coverage was one of the eight performance indicators on which disbursements were conditional. If three or more indicators were not met, the World Bank had the right to stop further disbursements.⁹² (Notably, in addition to DTP3 coverage, two other performance indicators had to do with immunization.) A 2007 World Bank loan to Argentina for the purposes of increasing the coverage of the Priority Public Health Programs

⁸⁷ The World Bank, "Loan Agreement (Essential Public Health Functions and Programs Project) Between Argentine Republic and International Bank for Reconstruction and Development," Loan 7412-AR, July 12, 2007, available at <http://www-wds.worldbank.org/external/default/2WOSContentServer/WDSP/LCR/2007/11/06/2D724C27D99ADACA8525732800735F6C/2_0/Rendered/INDEX/Loan07412/ARIconformed.txt>.

⁸⁸ International Monetary Fund (IMF), "Debt Relief Under the Heavily Indebted Poor Countries (HIPC) Initiative," March 31, 2011, available at <<http://www.imf.org/external/np/exr/facts/hipc.htm>>.

⁸⁹ Ibid.

⁹⁰ Joseph F. Naimoli, Deepti Tanuku, and Shilpa Challa. "HNP Brief #1—PRSPs, Immunization Targets, and Equity: Focus on Africa," *Health Nutrition and Population* (February 2005), available at <http://siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/Resources/281627-1109774792596/1_Immunization&Equity3Mar05.pdf>, 1.

⁹¹ World Bank, "Project Appraisal Document on a Proposed Adaptable Program Credit in the Amount of SDR 27.6 Million to the Republic of Bolivia in Support of the Second Phase of the Health Sector Reform Program," World Bank Doc. 22301-BO, June 12, 2001, available at <http://www-wds.worldbank.org/external/default/WDSPContentServer/WDSP/IB/2001/07/04/000094946_01062004021846/Rendered/PDF/multi0page.pdf>, 18.

⁹² Ibid., 33.

identified as “immunizations, tuberculosis, AIDS and STD, vector-borne diseases, emerging diseases, safe blood supply, emergencies and disasters, non-communicable diseases and risk factors, border health, and reproductive health, and any other program as the Bank may agree”⁹³ contained a clause that conditioned disbursements by the national government to each participating province on Public Health Activity Performance Indicators (PHA Performance Indicators): immunization coverage, disease control, and tuberculosis cases receiving directly observed treatment of tuberculosis (DOTs).⁹⁴ To receive the funds, provinces had to comply with PHA Performance Indicators and no amount was disbursed unless the national government verified, based on a report by an auditor, that PHS Performance Indicators set for the first 24 months of project implementation were met by the province. If there was non-compliance with *any* of the indicators, the amount allocated to a province was reduced on the basis of 25 percent reduction for each non-complied indicator.⁹⁵ Such arrangement effectively prioritizes the three benchmarked interventions over other health initiatives covered under the agreement.

Similarly, when Mauritania had to achieve a stated commitment of 70 percent DTP3 coverage by 2002, the Prime Minister himself followed the progress closely and visibly re-launched the immunization program for 2001 (the program had been languishing for the previous four years).⁹⁶ There was intense political pressure on program managers to achieve stated targets, “sometimes at all cost, and with few restrictions.”⁹⁷ What the Ministry of Health deemed necessary to achieve the coverage objective quickly was sanctioned by higher authorities: vaccination campaigns were initiated, “regions were instructed to reduce the dropout rate at the expense of vaccine wastage,” and ad hoc vaccination days were announced.⁹⁸

As these examples illustrate, conditioning funding on indicators prioritizes the interventions imbedded in the indicator, thus influencing the direction on national and provincial health policies. National health ministers may feel internal pressures from their colleagues in finance ministries as well because funding in other sectors may be dependent on the performance of the country’s immunization system. The widespread and broadly spanning use of immunization coverage indicators renders them a powerful tool of global governance and alters the dynamics of global governance by making indicator producers and users—WHO, the World Bank, MCC—“governors” of target national populations. The implications of this “governance at a distance” are discussed next.

⁹³ World Bank, “Loan Agreement (Essential Public Health Functions and Programs Project)” (n. 87), Appendix, para. 30.

⁹⁴ The latter refers to a component of a national strategy.

⁹⁵ If, during the subsequent six months, the province met the respective non-performing indicator, the amount was increased accordingly. World Bank, “Loan Agreement (Essential Public Health Functions and Programs Project)” (n. 87), schedule 2, para. 7(a).

⁹⁶ Joseph Naimoli et al., “HNP Brief #1”, 1.

⁹⁷ Ibid, 2.

⁹⁸ Ibid.

Immunization coverage indicators as a technology of governance

Governance, meant here as the means used to influence behavior, the production of resource, and the distribution of resources,⁹⁹ can take on different forms. Exercise of power and influence can occur through promulgation of legal rules and decisions, publication of scientific reports, military actions, transfer of funds, as well as through the collection, processing, and dissemination of information.¹⁰⁰

By collecting immunization data, processing it, and producing immunization coverage indicators, WHO influences states' immunization systems directly, by prompting engagement between its regional offices and states' ministries of health, and indirectly, by giving NGOs the basis on which to advocate for immunization campaigns and funders the reason to initiate immunization-related interventions for low-performing states and reward high-performing states (e.g., by providing funding for newly introduced vaccines).

The ecology within which the immunization indicators emerged—the development of WHO's institutional identity, WHO's relationship with donors and the World Bank, rising demands for credible indicators in development financing—transformed the indicators from a diagnostic tool to an intervention that impacts national health policies. The use of DTP3 coverage as a performance indicator in development financing by IFIs *de facto* directs states to prioritize immunization over other health initiatives. This not only impacts (positively) those individuals that benefit from increased focus on immunization but also potentially affects (negatively) those that would have benefited from non-vaccine interventions. In the case of the 2007 World Bank loan to Argentina, discussed above, this group would include, for example, persons with AIDS. Additionally, even the target group of immunization interventions—young children—may be disadvantaged by the absolute prioritization of vaccination. Consider, for example, diarrhea, which is a major cause of child mortality. Rotavirus vaccine is an intervention, but so are breastfeeding, oral rehydration therapy, supplementation with zinc and vitamin A, provision of antibiotics for dysentery, and basic water, sanitation, and hygiene (WASH). All have been proven effective. Many WASH interventions, for example, also reduce childhood growth stunting and thus provide indirect protection against several other causes of child mortality. In addition, although the vaccine intervention may be significantly cheaper, it tends to be substantially less effective for malnourished children with impaired immune responses or where vaccines are at high risk of contamination.¹⁰¹ In some

⁹⁹ Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 1), 78–80.

¹⁰⁰ Von Bogdandy and Goldmann, "Taming and Framing Indicators," this volume, pp. 52–85; Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 1), 78–80.

¹⁰¹ <<http://www.nature.com/nrgastro/journal/v8/n7/full/nrgastro.2011.103.html#B5>>; C.L.F. Walker, I.K. Friberg, N. Binkin, M. Young, N. Walker et al., "Scaling Up Diarrhea Prevention and Treatment Interventions: A Lives Saved Tool Analysis", *PLoS Med* 8(3), (2011): e1000428. doi:10.1371/journal.pmed.1000428; A.C. Serazin, L.A. Shackelton, C. Wilson, and M.K. Bhan, "Improving the Performance of Enteric Vaccines in the Developing World," *Nature Immunology* 11 (2010): 769–73.

countries, thus, children may benefit more from prioritizing non-vaccine anti-diarrheal interventions and nutrition programs, at least in the short term, than vaccine-based interventions.

Absolute prioritization of immunization may paradoxically compromise children's access to primary health care. Consider, for example, recent examples in the United States of pediatricians turning away children whose parents refused to vaccinate them.¹⁰² According to a 2001 American Academy of Pediatrics survey, 23 percent of physicians reported that they "always" or "sometimes" tell parents they can no longer be the child's pediatrician if they will not get the proper shots.¹⁰³ For countries whose aid is tied to high immunization rates, the risk of such behavior among health care providers is significantly higher, especially where health providers feel pressure from upper levels to produce high coverage rates.¹⁰⁴ This not only calls into question the propriety of using immunization coverage indicators as a proxy for health system performance, but also suggests that over-emphasis of immunization coverage indicators by IFIs sometimes can negatively affect children's access to health care.

Additionally, while indicators can highlight a problem, they can also de-humanize it. Although using DTP3 coverage may be a convenient way to assess the quality of a health care system, the indicator conceals the reality of children who have been refused health care because their parents refused to vaccinate or wanted to adhere to a vaccination schedule that was different from that mandated by health authorities, or of migrant and displaced children who are not included in the official counts because of their uncertain or disputed citizenship status and for whom health services are not available. The immunization coverage indicator also says nothing about children who get sick or die from non-communicable diseases because interventions that could help them may not get the same attention and priority as routine immunization.

Indicators, especially those produced by entities with scientific authority, can be more impenetrable than other acts of global governance because they can be perceived as *scientific*, *technical*, requiring *expertise*. When powerful peer institutions adopt the *scientific* indicator, their decisions acquire scientific authority as well. The wider the use of the indicator, the more legitimacy the indicator (and in turn, the

¹⁰² K. Fiore, "Strong Beliefs About Vaccines Work Both Ways," ABC News, March 4, 2011.

¹⁰³ Ibid.

¹⁰⁴ S.R. Onta, S. Sabroe, and E.H. Hansen, "The Quality of Immunization Data from Routine Primary Health Care Reports: A Case from Nepal," *Health Policy and Planning* 13:2 (1998): 131–9. In 1989, national immunization coverage for Nepal was reported as 95 percent for BCG, 80 percent for DPT and polio, and 69 percent for measles. The researchers found that the data in the register, the primary health care (PHC) report, which is the next level, and the District Health Office (DHO) report, the next higher level, were all different and the number of all vaccine types reported from the PHCs outlets was higher than the number of vaccines in the immunization register. The difference ranged from 15 percent for BCG vaccine to 174 percent for polio vaccine. The cumulative average difference between the immunization register and the DHO report was 64.2. The researchers attribute the measurement error to top-down target-setting, lack of participation by the lower level health providers in the process, and performance incentives.

indicator producer) receives. Through this iterative process, the indicator becomes taken for granted. When that happens, decisions that are based on the indicators may be less likely to be questioned as well. In that circular process, the people behind the numbers get lost and the gap between the indicator-driven policies and the needs and rights of these people may widen.

Individuals affected by the policies of international organizations may be able to mediate the organizations' exercise of power through contestation. For example, a transnational group of workers' representatives (spearheaded by the International Trade Union Confederation) successfully campaigned for the World Bank to drop its Employing Workers indicator on the grounds that it was biased in favor of labor market deregulation and compromised workers' rights. However, in the case of immunization coverage indicators there has been virtually no public contestation either of production or of far-reaching use of the indicator.¹⁰⁵ Lack of contestation is not due to the infallible nature of the indicators: the 2011 World Health Statistics report itself contains a paragraph-long disclaimer, including a statement that "because of the weakness of the underlying empirical data in many countries, a number of the indicators are associated with significant uncertainty."¹⁰⁶ A number of possible explanations for the lack of contestation exist.

First, it is possible that, unlike in the case of the Employing Workers indicator, there is a general agreement on the theory of change imbedded in the indicator. That is, there is widespread acceptance that immunization is the most efficient intervention in reducing child mortality and morbidity. This understanding might be challenged if vaccine-preventable diseases were not the leading (or even a major) cause of child mortality and morbidity. As discussed above, categories of vaccine-preventable and other diseases are fluid, yet they are often presented as factual categories. Individuals without the specialized knowledge of what diseases cause child mortality and how they are classified may find it difficult to assess the link between access to vaccines (versus other interventions) and reduction in child mortality. The role of *scientific expertise* and *scientific authority* in circumscribing contestation of indicators is discussed in more detail below.

Second, negotiation over the construction of indicators between states and WHO may be taking place privately, in closed-door meetings or through informal communication. The fact that WHO publishes the data reported from states alongside its estimates incentivizes all parties to ensure that the numbers strongly

¹⁰⁵ There is anecdotal evidence that WHO faced pressure from countries to make its methodology for producing immunization coverage estimates clearer, replicable, and publicly available. A representative from WHO mentioned that a revised methodology should be published by the end of 2010. Interview with Interviewee 1. The 2011 World Statistics Report contains a paragraph in which WHO states that "[i]t is WHO policy on statistical transparency to make available to users the methods of estimation and the margins of uncertainty for relevant indicators. However, because of space restrictions, printed versions of the World Health Statistics series include uncertainty ranges for only a few indicators. Further information on the margins of uncertainty for additional indicators will be made available at the Global Health Observatory web site." WHO, *World Health Statistics: 2011* (Geneva: WHO, 2011), available at <http://www.who.int/whosis/whostat/EN_WHS2011_Full.pdf>, 8. At the time of writing, no additional information was yet available.

¹⁰⁶ WHO, *World Health Statistics: 2011* (n. 105), 8.

correlate. Indeed, during the production of indicators, WHO provides states with the opportunity to engage in a dialogue if they disagree with the immunization coverage indicators by communicating the indicators to the Ministry of Health prior to their publication.¹⁰⁷ A state has approximately four to six weeks to react to the estimates, may challenge the numbers as inaccurate, and if it “makes a good case for why certain things need to be taken into consideration, [WHO] will take the concerns into consideration and revise the estimates.”¹⁰⁸ (No aspect of the WHO’s interaction with the state is revealed publicly, including the standards used to determine whether a country’s challenge amounts to “a good case.”) There also appears to be at least some room for outside engagement with WHO during the process of indicator production: WHO notes that it consults with local experts and attempts “to put the data in the context of local events, both those occurring in the immunization system . . . and more widely occurring events.”¹⁰⁹ In addition, over the decade prior to 2011 reportedly four external and two internal reviews of immunization indicators took place.¹¹⁰ In 2011, WHO noted that its methods and findings for the 2010 report “were reviewed by a group of external experts with broad experience in immunization systems and survey methodology.”¹¹¹ The names of the reviewers, the methodology of the review, or the results of the reviews are not publicly available. It is not clear who constitutes an “expert” for purposes of these engagements and whether “experts” include NGOs, community leaders, or local health providers.

Third, it is a lot more difficult and costly for a state to sustain a high level of performance on many facets of its health care system than on one discrete area. For this reason, states may prefer that IFIs use the immunization coverage indicator as a proxy for the overall performance of the health system. DTP3 coverage is an “easy” target for compliance: (a) because states have been reporting on immunization coverage for decades, most states have an existing infrastructure not only for the provision of the vaccine but for monitoring and reporting of the coverage (those countries who report below-threshold coverage can be eligible for funds from GAVI to establish or improve monitoring and reporting infrastructure); and (b) because states already report immunization coverage data to WHO, states do not have to

¹⁰⁷ Most of the dialogue takes place electronically, although on rare occasions telephone conversations may take place with national program managers and from time-to-time face-to-face meetings will take place where WHO representatives will discuss the methods of indicator generation with country’s representatives. Interview with Interviewee 1. Occasionally, countries are invited to meet with WHO and UNICEF and PAHO to review the country’s data and understand the process WHO/UNICEF adopts to arrive at the estimates, but what exactly is discussed at these meetings is not publicly revealed. Interview with Interviewee 2. In any event, these kinds of meetings seem to be designed to aid countries in understanding the estimation process. Interview with Interviewee 1; Interviewee 2.

¹⁰⁸ Interview with Interviewee 1.

¹⁰⁹ WHO, “WHO Vaccine-Preventable Diseases” (n. 61), 14.

¹¹⁰ Interview with Interviewee 1. Reviewers reportedly included former WHO Assistant Director General, World Bank representatives, an independent statistician who had experience with immunization coverage, the head of research from DHS surveys, president of Chinese Academy of Preventative medicine, professor of biology from Penn State University, professor of pediatrics, a representative from UNFPA, and representative from Statistics Norway.

¹¹¹ WHO, “WHO Vaccine-Preventable Diseases” (n. 61), 15.

expend additional resources in order to monitor and report on the target. States may also prefer the imperfection of the immunization coverage indicator to the political cost of a comprehensive health system indicator (see discussion of WHO-produced 2000 health system ranks above).

Fourth, the cost of contesting the indicators might be prohibitive for a state with limited resources. For states that have poor data collection services or that prefer to allocate limited resources to other areas of public health, having WHO produce estimates based on surveys or other data may be preferable even at the expense of living with some degree of uncertainty. The state may also perceive itself to be lacking the necessary knowledge or expertise to contest WHO's indicators because indicators generally, and immunization coverage indicators in particular, are often perceived as products of "*technical*" and "*scientific*" process.¹¹²

At first glance, immunization coverage indicators are simple. Unlike the case of health system rankings, which require selection and measurements of several categories and provide opportunities for challenges to the categories selected, the immunization coverage indicator for a particular antigen is concerned with one measurement and one category—number of administered vaccines for a particular antigen. On the other hand, technical difficulties may arise during the data collection and reporting process. For example, it may not be easy to arrive at the denominator—census data may not be accurate or not available, there are questions as to who should be included (e.g., should children of migrant workers be included in the target population and, if so, of what country?), or the age of the children may not be easily established. There is evidence that some health care providers do not know how to compute immunization rates.¹¹³ It may also be difficult to ascertain which dose of a multi-dose vaccine a child is receiving—records may be poor or unavailable and the caregiver's recall may be inaccurate. Surveys are a good supplementary means for assessing immunization coverage, but they are expensive and NGOs or local community groups may not possess the technical knowledge or the resources to conduct them.

WHO embraces and encourages the perception that the production of the indicators requires technical expertise which many countries lack. Paradoxically, the organization's failure to articulate clearly and make publicly available its methodology for producing the indicators may add to the image of the indicators as *technical* and *scientific*. What makes the paradox possible is the scientific authority, which, as discussed above, WHO has acquired through purposeful efforts over the years—as an expert in global health statistics, the organization is uniquely positioned to collect global data on immunization (among other aspects of health care) and engage in the technical process of converting multi-sourced data into easily understood indicators. An indicator that is produced through this process acquires credibility that may be impenetrable to challenge except by those with

¹¹² Davis, Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 1), 86–9.

¹¹³ João Carlos de Timóteo Mavimbe, et al., "Immunization Coverage in Mozambique: From Concepts to Decision-making," *Health Policy* 79 (2006): 95 (noting that few health district directors understood how immunization coverage rates were computed).

similar scientific expertise (who can contest the quality of data or methods of production) and those with intimate knowledge of institutional history (who can challenge the producer's authority). As Daniel Esty notes then, "when a matter is largely scientific or technical, having designated supranational experts address the problem may be uncontroversial";¹¹⁴ at the same time such delegation shifts the balance of power toward "technical" experts.¹¹⁵

Indeed, the main contestation of immunization indicators appears to be driven by members of the scientific community. Most of the critique, voiced through scientific studies published in science journals, is focused on the *process* by which the immunization coverage indicators are produced and the quality and propriety of the underlying data (i.e., *content* of the indicators). Many of the studies are funded by development agencies, often in partnership with WHO or by WHO itself.

Fifth, using the indicators produced by an entity that is widely perceived as an expert in health statistics, lends scientific authority to the users of indicators. IFI's decisions based on indicators may thus be perceived as more objective and legitimate. In domestic settings, when scientific advisory committees are established to assist policymakers with decision-making, science inevitably becomes politicized, with policymakers recruiting scientists to legitimate their policy choice.¹¹⁶ In the global arena, WHO is the main scientific expert; there is no alternative scientific authority that can delegitimize decision based on WHO's data.

Since the 1990s, the only significant challenge to WHO's *scientific authority* to produce indicators has come from former WHO officials.¹¹⁷ In 2004, Christopher Murray, who had previously served as WHO's Executive Director of the Evidence and Information for Policy Cluster publicly questioned WHO's ability to provide effective global reporting and recommended the establishment of a new global health monitoring organization. Murray claimed that all WHO indicators (not just immunization coverage estimates) are subject to intense political pressure, making

¹¹⁴ Daniel C. Esty, "Good Governance at the Supranational Scale: Globalizing Administrative Law," *Yale LJ*, 115, 1490 (2006): 1511.

¹¹⁵ Davis Kingsbury, and Merry, "Indicators as a Technology of Global Governance" (n. 1).

¹¹⁶ Peter Weingart, "Scientific Expertise and Political Accountability: Paradoxes of Science in Politics," *Science and Public Policy* 26, (1999): 151–61.

¹¹⁷ Limited challenges to WHO's authority to produce indicators should be distinguished from challenges of WHO's governance more broadly, of which there are many. As a most recent example, at the 2011 World Health Assembly meeting in Geneva, Corporate Accountability International, which represents 100 organizations from 24 countries, claimed that WHO is compromising its independence and mission of improving global health and delivered a letter to WHO Director General Margaret Chan asking that the UN agency reject corporate influence and maintain its independence. Similarly, Council of Canadians contended that WHO's policy on water has favored the interest of corporations who seek to privatize many public water sources. Tom Paulson, "Critics Say World Health Organization Too Cozy with Corporate Interests," KPLU, March 19, 2011, available at <<http://humanosphere.kplu.org/2011/05/critics-say-world-health-organization-too-cozy-with-corporate-interests/>>. In May 2011, a number of high-profile NGOs issued a statement expressing concerns about governance of the World Health Organization (WHO) and other public health bodies, as regards the management of conflicts of interest, in particular the organization's relationship with corporate interests. James Love, "NGO Statement on WHO Governance and the Management of Conflicts of Interest and the Future of Financing for WHO," *Knowledge Ecology International*, May 18, 2011, available at <<http://www.keionline.org/node/1135>>.

them “ill suited for the role of global monitoring and evaluation of health.”¹¹⁸ In July 2007, Murray became the Executive Director of the Institute for Health Metrics and Evaluation (IHME), based in the University of Washington, established to “monitor global health conditions and health systems, as well as evaluate interventions, initiatives, and reforms and as such provide high-quality and timely information on health so that policymakers, researchers, donors, practitioners, local decision-makers, and others can better allocate limited resources to achieve optimal results.”¹¹⁹ The IHME’s Board is filled with former WHO “insiders”: a former Executive Director of Evidence and Information for Policy, a former Special Envoy of the WHO Director General in Human Resources for Health and a Board Chair of the Global Health Workforce Alliance, a former consultant to WHO, a former WHO Executive Board Member, a former member of WHO expert panel, a former technical advisor in WHO’s Department of Communicable Disease Surveillance and Response, and a former director of WHO’s Division of Family Health (Gro Harlem Brundtland, Murray’s former boss at WHO, was on the first Board of the Institute).

In 2008, a study conducted by Stephen S. Lim et al. under the auspices of IHME¹²⁰ and funded by the Bill and Melinda Gates Foundation challenged GAVI’s performance-oriented Immunisation Services Support (ISS) Program, which effectively provides to countries \$20 for each additional child above target who receives the third dose of DTP vaccine.¹²¹ Lim has worked alongside Murray at the Evidence and Information for Policy Cluster at the World Health Organization (WHO) and also acted as a reviewer for the WHO/UNICEF Estimates of National Immunization Coverage. The authors of the study argued that performance-oriented funding contributes to over-reporting of immunization coverage levels: in all 193 countries the study analyzed, officially reported estimates and administrative data exhibited greater variation over time than survey data, and some countries receiving ISS payments significantly over-reported DTP3 coverage. The discrepancy between country-reported coverage rates, WHO indicators, and coverage rates produced by surveys was quite stark: 1990 official reports estimated global DTP3 coverage at 83 percent, WHO at 75 percent, survey-based estimates at 65 percent; from 1999 to 2006, officially reported estimates showed a 9 percent increase in DTP3 coverage (from 81 percent to 90 percent), WHO an 8 percent increase (from 71 percent to 79 percent), and surveys showing an increase of

¹¹⁸ Christopher J.L. Murray, Alan D. Lopez, and Suwit Wibulpolprasert, “Monitoring Global Health: Time for New Solutions,” *British Medical Journal* 329 (2004): 1096–100. For WHO’s response, see Kenji Shibuya, Suzanne Scheele, and Ties Boerma, “Health Statistics: Time to Get Serious,” *Bulletin of the World Health Organization* 83 (2005): 722.

¹¹⁹ Christopher J.L. Murray, “Director’s Statement,” *Institute for Health Metrics and Evaluation (IHME)*, available at <<http://www.healthmetricsandevaluation.org/who/director.html>>.

¹²⁰ Stephen S. Lim et al., “Tracking Progress Towards Universal Childhood Immunization and the Impact of Global Initiatives: A Systematic Analysis of Three-Dose Diphtheria, Tetanus, and Pertussis Immunization Coverage,” *The Lancet* 372 (2008): 2031–46.

¹²¹ GAVI Alliance, “Immunisation Services Support,” available at <<http://www.gavialliance.org/support/what/iss/index.php>>.

4.9 percent.¹²² The study also found that in 2006, compared with survey-based estimates, official reports overestimated the number of children immunized with DTP3 by 20.2 million and WHO overestimated the number by 6.1 million.¹²³

Controversy over the study ensued not only because it challenged a core GAVI program but because it risked undermining the credibility of WHO's immunization coverage indicators and provided support to Murray's earlier challenge to the organization's authority to produce them. WHO published a response in *Lancet*, challenging the methodology of the Lin study,¹²⁴ but also using it as an opportunity to call for improvement in (and additional funds for) local administration, monitoring, and data-reporting capacities. The debate appears to have subsided; WHO continues to produce immunization coverage indicators and Murray's center continues to produce health statistics, although it does not appear to have acquired the status of a real competitor to WHO.

Conclusion

Given WHO's institutional interest in having its indicators broadly adopted, without real challenge to WHO's scientific authority, decisions based on WHO's indicators are legitimated through a recursive process between the producer and users. Through this process, over time, the use of the indicators is settled so that it becomes taken for granted.¹²⁵ When this happens, there is a risk that people affected by the indicator-driven policies will lose the opportunity to have their interests taken into consideration in the decision-making processes, which in turn may lead a gap between the needs and rights of the affected people and the policies directed at them.

Some of the risk may be mitigated through the already-existing legal and regulatory channels. Individuals affected by the indicator-driven policies could use national courts to challenge the policies themselves, for example, on constitutional grounds. This option, however, may not be apparent to those who could have benefited from de-prioritized medical interventions, as they may not even be aware of the de-prioritization. Governments can in turn claim lack of resources to support a particular intervention without revealing that another intervention is being prioritized as a result of agreements with IFIs or, conversely, may claim that their hands are tied by IFIs. Affected populations may resort to internal accountability mechanisms that some IFIs have, for example, by bringing a grievance to the World Bank Inspection Panel. This, however, would require individuals to understand how the indicators are produced and what the better alternatives might be,

¹²² Lim et al., "Tracking Progress Towards Universal Childhood Immunization" (n. 120), 2040.

¹²³ Ibid.

¹²⁴ Tony Burton et al., "Measurement of Immunisation Coverage," *The Lancet* 373 (2009): 210–11.

¹²⁵ T.C. Halliday, "Recursivity of Global Norm-Making: The Sociolegal Agenda," *Annual Review of Law and Social Science* 5 (2009): 263–89.

something that might be difficult to achieve for those without the required *expertise*, as discussed above.

GAL principles, such as transparency, participation, reason-giving, and review, might alleviate some of the concerns surrounding the use of indicators.¹²⁶ GAL principles can be applied at three distinct points in time: production, promulgation, and use of indicators. If populations affected by indicator-driven policies are provided with sufficient information about how the indicator is produced and why it is used (as compared to other alternatives), and are given the opportunity to comment both on the production and on the effects of using, e.g., the DTP3 indicator as a performance measure for other health initiatives, it may be possible to reduce (or even eliminate) any gap between the needs and rights of populations and indicator-driven policies. In the context of immunization coverage indicators, applying GAL to *production* of indicators might mean ensuring that the data is collected, analyzed, and triangulated, that the methodology is replicable and publicly available in a form that is accessible by those outside the scientific community. The process of indicator production should be subject to regular independent review that not only ensures quality of data and integrity of the process, but that also identifies limitations of the indicators. Producers could then be required to disclose limitations or uncertainties associated with the use of the indicator as well as to reveal the imbedded standards and theories in the same way that, for example, sellers of securities are required by the US anti-fraud regime to disclose all material information related to those securities (including Risk Factors associated with purchase of the securities).

Promulgators of indicators could be required to ensure that publication of indicators is subject to notice and comment procedures and that they are accompanied by summary narrative explaining the purposes for which the indicators are intended so as to overcome any deficit of accessibility caused by the real or perceived *technical* nature of indicators. Both producers and promulgators could also be required to consider and disclose the unintended but foreseeable uses of indicators, risks associated with the use of indicators and, where applicable, issue warnings against the specific uses of the indicators.

Such an approach is not without precedent. Industry groups, such as the College Board, a not-for-profit membership association composed of more than 5,700 schools, colleges, universities, and other educational organizations, issue guidelines on the Uses of College Board Test Scores and Related Data, not just for different users (itself, Institutions, Agencies, and Organizations) but also for different *uses* (counseling, recruiting and marketing, admission and placement, and credit).¹²⁷

¹²⁶ B. Kingsbury, N. Krisch, and R.B. Stewart, "The Emergence of Global Administrative Law," *Law & Contemporary Problems* 68 (2005): 15.

¹²⁷ For example, users of test scores for placement are told to "[p]ublicize their placement and credit policies, making clear which tests are used, how students gain access to them, what scores are required. . .," "[e]stablish standards regarding the age of scores and norms used for interpretation when setting placement and credit policies," and "[p]eriodically validate examinations and standards used for placement and credit to ensure their continuing relevance." <<http://professionals.college-board.com/profdownload/guidelines-on-uses-of-college-board-test-scores-and-data.pdf>>.

Similarly, collaboration of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education produced "The Standards for Educational and Psychological Testing," which address test construction, evaluation, and documentation, as well as rights and responsibilities of test takers and test users. And, of course, product liability regimes and rules surrounding advertising and marketing products require manufacturers to disclose risks associated with the product's use, including risks associated with unintended but foreseeable use.

GAL principles could also enhance the indicator-based decision-making process. For example, at the World Bank and the International Monetary Fund, ex ante conditions (i.e., compliance with performance indicators) are normally required to be completed before a financing arrangement is presented to the respective Executive Boards for approval. This means that actions required to comply with the indicators are only formally brought to the attention of the Executive Directors once conditions have been complied with.¹²⁸ As a result, "[t]he design of prior actions . . . is not reviewed officially and discussions on policy and institutional reforms are undertaken largely between staff of the [international financial institutions] and the country authorities."¹²⁹ This process circumscribes the opportunity for contesting the use of particular indicators as ex ante conditions and increases leverage of international financial institutions "vis-a-vis the governments of developing countries."¹³⁰ Making the design of performance indicators reviewable and the results of the review available to the public, for example, could open the use of particular indicators to a broader debate in the course of which other alternatives could be identified. Identification of a plausible alternative to the DTP3 indicator may in itself prompt questioning of whether DTP3 is being used appropriately and how the effects of using DTP3 compare with the effects of using the alternatives.

¹²⁸ Tan, *Governance through Development* (n. 83).

¹²⁹ Ibid.

¹³⁰ Sarah Babb and Ariel Buira, "Mission Creep, Mission Push and Discretion in Sociological Perspective: the Case of IMF Conditionality," paper presented at the XVII G24 Technical Group Meeting, March 8–9, 2004, Geneva.

PART III

TRANSLATION,
TRANSPLANTATION, AND
ADAPTATION: THE
RELATIONSHIP BETWEEN
“GLOBAL” AND “LOCAL”
IN INDICATOR PRODUCTION
AND USE

This page intentionally left blank

Internally Displaced Population in Colombia

A Case Study on the Domestic Aspects of Indicators as Technologies of Global Governance

*René Uruña**

The notion of indicators as technologies of global governance, as advanced in this volume, denotes at least two parties: one “producing” the indicator, the other being measured by it. Each of these parties may be, in turn, unpacked so as to include competing actors and agendas on each side of the equation. Thus, research of the challenges of this technology requires two complementary paths. First, an exploration of the producing end of the equation: How are indicators created? By whom? Do they enhance or reduce accountability? Can they be regulated? Part II of this book explores these questions. A second area of exploration is the “receiving” end of the indicator; that is, research focusing on those whose performance is being measured and compared.

This chapter belongs to that second area. It presents a case study on the role of international indicators in the implementation of internally displaced population (IDP) policy in Colombia, and argues that monitoring is only part of the indicators story, as there is a dialectic element to the process of creating and applying them. Focusing on the role of indicators in domestic politics, and thus sharing some of the preoccupations featured in this volume by Stone’s study on the “Problems of Power in the Design of Indicators of Safety and Justice in the Global South,” and by Hagan and Zaloznaya’s work on Belarus, this chapter argues that indicators are also creatively used for local purposes, not necessarily connected with their original, global origin. Both governments and non-state actors adopt indicators as part of their rhetoric. Moreover, indicators become a variable in the interaction between independent branches of domestic power. It is also likely that those who are measured will try to influence the scale of measurement, making indicators an important aspect of their strategic political choices. Ultimately, the case study

* Drafts of this chapter were presented at the conference “Midiendo el Derecho: El Uso de Indicadores en el Ejercicio Poder Global” (Bogotá, April 2010) and “Indicators as Technologies of Global Governance” (New York, September 2010), both of which were supported by the Carnegie Corporation of New York. I am grateful to Ryan Goodman and the students of his Fall 2010 International Law seminar at NYU for comments to an earlier draft. The research assistance of Sebastián Machado is gratefully acknowledged. The usual caveat applies.

featured in this chapter suggests that governance through indicators is more complex than a one-way, top-to-bottom process. On the contrary: there is a process in which an indicator, once it enters local debates, seems to gain a life of its own. It is by looking at the intimacies of its domestic life that we get a better sense of the inner working of this technology of governance, the reasons behind its success, and the specific challenges it poses.

This chapter is divided into four sections. The first argues that IDP policy is a species of global governance, and introduces the role that indicators play therein. The second section is a brief introduction to the human tragedy of internal displacement in Colombia, and sketches the domestic institutional reaction to it. That section also features the role that indicators have played in domestic IDP policy in Colombia. The third section features the domestic aspects of indicators as technologies of global governance, and explores the mechanisms of creative use of global IDP indicators in the Colombian context and consequences for domestic policymaking. It does so with reference to four axes: (a) domestic IDP indicators as input for global decisions; (b) the role of the news media in the turn to indicators as technologies of global governance; (c) IDP indicators as social mobilization; and, finally, (d) IDP indicators and their role in the interaction among separate branches of power. The fourth and final section concludes.

IDP indicators and global governance

IDPs came into the global spotlight as successive humanitarian crises forced massive populations to move within the borders of single states, mainly since the early 1990s. To be sure, forced displacement had been in the agenda of international institutions for decades. The League of Nations, for instance, had a High Commission for Refugees, created in 1921 under the direction of Fridtjof Nansen (who previously led the repatriation of prisoners of war from Siberia, acting as High Commissioner for the League of Nations¹). Seeking to address the problem of the exiled population fleeing the Bolshevik regime after 1917, Nansen proposed and implemented the so-called "Nansen Passport," that is, an identity document (yet not a passport, *stricto sensu*) issued by adherent states, valid for a year, allowing the bearer to return to the country issuing it.² The "passport" was first issued to Russians, but was then extended to Armenians in 1924, and then to Kurds, Turks, Assyrians, and Syrians in 1928.³

The Nansen initiative is an early example of what would come to be the default approach to the problem of displaced population in most of the twentieth century. In essence, the approach was that forceful displacement became a problem as

¹ Louise W. Holborn, "The League of Nations and the Refugee Problem," *Annals of the American Academy of Political and Social Science* 203 (1939): 124.

² Louise W. Holborn, "The Legal Status of Political Refugees 1920–1939," *American Journal of International Law* 32 (1938): 684.

³ Laura Barnett, "Global Governance and the Evolution of the International Refugee Regime," *International Journal of Refugee Law* 14 (2002): 242.

populations or individuals crossed borders—thus becoming, for example, “refugees,” or asylum seekers. Displacement was an inter-governmental problem, which concerned the relation between states.⁴ This premise was confirmed by the effects of World War II, as the allies established the United Nations Relief and Reconstruction Agency (UNRRA) in 1944. UNRRA existed until 1947, when its mandate ended. In 1948, a temporary International Refugee Organization was set up as an agency of the United Nations (UN). Soon after, though, it became evident that the refugee problem was not of a temporary nature, and a permanent UN High Commissioner for Refugees (UNHCR) was created in 1950. Like its predecessors, and due to the demands of the particular moment in history in which it acted, the UNHCR was unconcerned with population displaced within a country. In fact, the agency lacked a specific mandate to deal with such populations under its Statute.⁵

IDP policy as global governance

Article 9 of the UNHCR Statute, though, did allow for the High Commissioner to “engage in such additional activities, including repatriation and resettlement, as the General Assembly may determine, within the limits of the resources placed at his disposal.” As the tragedy of internally displaced people began to be evident, mainly outside Europe, the alternative offered by Article 9 proved useful. Thus, in the context of the Sudanese crisis of the early 1970s, the UNGA “urged the organizations associated with the United Nations and all Government to render the maximum possible assistance to the Government of Sudan in the relief, rehabilitation of Sudanese refugees coming from abroad *and other displaced persons*.”⁶ Since then, the UNHCR has seen its mandate with regards to IDPs become broader and broader until 1997, when the now default formulation was established, which gave competence to the UNHCR to deal with the issue.

Despite the interest of the UNHCR as the organ of an international organization, IDPs are also (and perhaps, predominantly so) a domestic problem: IDPs are protected by domestic laws, and often are displaced by internal armed conflict, according to the 1948 Geneva Conventions. IDPs affect distribution of wealth, land ownership, and gender and ethnic victimization, all within a single state. Ultimately, IDPs are first and foremost a responsibility of the state within which the displacement occurs. But, as we have seen, the UNHCR has much to do and say about the problem. Reaction to the IDP challenge is, therefore, a point of contact between the agenda of an organ of a traditional inter-governmental organization (the UNHCR) with the agenda of national governments, their interests, and

⁴ In the same sense, see Holborn, “The Legal Status of Political Refugees 1920–1939,” 387. For a useful historical review, see Barnett, “Global Governance and the Evolution of the International Refugee Regime” (n. 3), 239–45.

⁵ Statute of the Office of the United Nations High Commissioner for the Refugees, UN General Assembly (GA) Res. 428 (V) of 1950, UN Doc. A/1775 (1950).

⁶ UNGA Res. 2958 (XXVII), December 12, 1972, UN Doc. A/8918 (1972), para. 3 (emphasis added).

those of other national power structures, a contact that is often bound to become a clash, as internal displacement becomes a sensible part of domestic politics, or is even caused by the very government primarily responsible for the victims.

The role of international law in IDP policy is a reflection of this circumstance. It has become common to argue that soft law plays an important role in the context of IDPs.⁷ This all-important role of soft instruments can be explained by the middle ground between international and domestic politics where IDPs stand. Consider the central normative piece to be found in IDP policy: the Guiding Principles on Internal Displacement, issued by the UN's Secretary-General's Special Representative on IDPs.⁸ The legal status of the Principles is rather ambivalent, considering that it is not a UN declaration, nor is it an attempt at codifying customary international law.⁹ Rather, it is a study of domestic legislation and analogous regulation (e.g., refugee law) which is, in the words of the Representative of the UN Secretary General for Internal Displacement, "consistent with international law."¹⁰ The Guiding Principles and their accompanying legal annotations make no reference to the Colombian situation, although they do rely twice on Colombian law as state practice in support of two principles therein enshrined: the right not to be discriminated against,¹¹ and the right to have access to an education.¹²

At the heart of these principles is the realization that several matters affecting IDPs are indeed covered by traditional (hard) international instruments, such as the right to life in human rights treaties, or the principle of distinction in international humanitarian law. Such is the international aspect of the problem. However, there are other matters affecting IDPs that concern mainly domestic jurisdictions, for example, the problem of identification within the state, compensation for property or land lost during the displacement, or the possibility of finding a safe place within one's own state. Facing such a situation, norm entrepreneurs (e.g., activist and academic networks, and the UN Representative of the Secretary-General for IDPs) quickly rose to support the drafting and adoption of some sort of norm that would address the limitations of the international aspect of the problem.¹³ The answer was the Guiding Principles.¹⁴ While no hard international instrument was available and challenges fell under each state's sovereignty, some degree of governance could still

⁷ For a recent example, see Phil Orchard, "Protection of Internally Displaced Persons: Soft Law as a Norm-Generating Mechanism," *Review of International Studies* 36 (2010): 281.

⁸ Available at <<http://www.idpguidingprinciples.org/>>. The principles were recognized by the GA as "an important international framework for the protection of internally displaced persons." GA Res. 60/L.1, UN Doc. A/60/L.1, para. 132.

⁹ See Walter Kälin, "Guiding Principles on Internal Displacement—Legal Annotations," ASIL-Brookings Institution/University of Bern Project on Internal Displacement, Occasional Paper No. 38, 2008, 6.

¹⁰ Ibid.

¹¹ Ibid., 12.

¹² Ibid., 108.

¹³ See Orchard, "Protection of Internally Displaced Persons: Soft Law as a Norm-Generating Mechanism" (n. 7).

¹⁴ On the role of international norm entrepreneurs in the IDP context, see Roberta Cohen and Francis M. Deng, *Masses in Flight: The Global Crisis of Internal Displacement* (Washington: Brookings Institution Press, 1998), 283–5.

be exercised through the principles. IDP governance, then, is not strictly national or international, but seems to include several layers of domestic governance, which is complemented by the international action of the UN and several networks of activists. A certain balance was, in this way, struck: while the primary responsibility still fell on states, international involvement remained possible. Francis Mading Deng, Representative of the UN Secretary-General on IDPs, clearly portrays this dynamics in his description of his work at the UN:

In my dialogue with governments—one of the requirements of my mandate—the first five minutes with the head of state is crucial to assure them of my recognition of the problem as internal and therefore under state responsibility. Having emphasized my respect for their sovereignty, I quickly move on to present the positive interpretation of sovereignty and the supportive role of international cooperation. Once I establish a cordial climate, candid and constructive dialogue can follow with little or no constraint in the name of sovereignty.¹⁵

And then Deng concludes:

the critical issue becomes how the international community can intercede to overcome the obstacles of negative sovereignty and ensure access for the needy population.¹⁶

Such is indeed the challenge. IDP policy is a combination of binding and non-binding instruments that is developed, implemented, and enforced by public, private, and semi-public agencies of national and international origin. It is a paradigmatic example of a global regulatory regime.¹⁷ IDPs are a primarily domestic issue, and the reasons behind any international action are not presumed, but need to be carefully spelled out. In the case of the UNHCR, GA Resolutions 47/105 (1992), 48/116 (1993), and 49/169 (1994) put forward the basic requirement for the agency's involvement in an IDP situation, including the specific request from the Secretary-General or other competent UN organ, and the consent of the state concerned. This need for justification points to a larger issue. Both global IDP policy in general, and the UNHCR in particular, have been perceived as too selective and too unpredictable in their approach to domestic crisis: why get involved in this context and not in that one? Why that action and not the other? Why this agency, here?¹⁸ Ultimately, having one foot in international politics and one foot in domestic politics, global IDP policy is in the constant need of proving itself rational and predictable, both to its international principals and to its domestic addressees.

¹⁵ Francis M. Deng, "The Global Challenge of Internal Displacement," *Washington University Journal of Law and Policy* 5 (2001): 145.

¹⁶ *Ibid.*, 145.

¹⁷ Previous scholarship has made this point with regards to other aspects of the UNHCR's mandate. See, for example, B.S. Chimni, "Co-option and Resistance: Two Faces of Global Administrative Law," *New York University Journal of International Law and Politics* 37 (2005): 799; Mark Pallis, "The Operation of the UNHCR's Accountability Mechanisms," *New York University Journal of International Law and Politics* 37 (2005): 869.

¹⁸ In what counts as strong wording for the General Assembly, GA Res. 58/177 of December 22, 2003 "Emphasizes the need to strengthen further inter-agency arrangements and the capacities of the United Nations agencies and other relevant actors to meet the immense humanitarian challenges of internal displacement, and underlines in this regard the importance of an effective, accountable and predictable collaborative approach," UN Docs A/RES/58/177, para. 13.

Indicators enter the scene

Part of the reaction to this call was the adoption of a "cluster approach" in 2005, in which several agencies concerned with humanitarian action, all belonging to the Inter-Agency Standing Committee (IASC),¹⁹ agreed to take on responsibilities according to their expertise, both at the global and country levels.²⁰ In essence, the effort in the cluster approach seeks, at the global level, "to strengthen system-wide preparedness and technical capacity to respond to humanitarian emergencies."²¹ At the country level the goal is, in turn, to "to strengthen humanitarian response by demanding high standards of predictability, accountability and partnership in all sectors or areas of activity."²²

Due to its expertise, the UNHCR took most of the responsibility concerning IDPs.²³ And it turned to indicators in order to achieve its aims. In the clusters where the UNHCR is the lead agency, a target objective is the development and implementation of some kind of indicator.²⁴ Thus, in the Camp Coordination and Camp Management (CCCM) Cluster, the UNHCR seeks to "develop and disseminate appropriate standards and indicators for CCCM,"²⁵ in the Emergency Shelter Cluster, the objective is to "establish impact indicators which will allow proper monitoring."²⁶ More indirectly, perhaps, in the Protection Cluster, the goal is more generally to "strengthen information management."²⁷

This move to indicators is, in fact, a reflection of a general trend within the UNHCR. In 2005, the UNHCR Camp Indicator Report (CIR) was replaced by the Standards and Indicators Report (SIR) in order, according to the agency, "to further the implementation of results-based Management (RBM) and facilitate multi-year planning based on measurable indicators illustrating the well-being of the population of concern."²⁸ As a result, quantitative reporting and monitoring became a central aspect of the architecture of IDP governance,²⁹ and of its local reception and use.

¹⁹ The IASC was created in 1992, following the mandate of GA Res. 46/182 of December 19, 1991 (UN Docs A/RES/46/182, para. 38). In practice, it is an informal forum where several UN and non-UN humanitarian agencies meet to coordinate their action. See <<http://www.humanitarianinfo.org/iasc>>.

²⁰ On the cluster approach, see IASC, "Guidance Note on Using the Cluster Approach to Strengthen Humanitarian Response," November 24, 2006, available at <<http://www.reliefweb.int/humanitarianreform/IASC%20GUIDANCE%20NOTE%20ON%20CLUSTER%20APPROACH.pdf>>.

²¹ *Ibid.*, 2.

²² *Ibid.*

²³ See IASC, "Guidance Note on Using the Cluster Approach to Strengthen Humanitarian Response" (n. 20), 3.

²⁴ See UNCHR, "UNHCR's Contribution to the Inter-agency Response to IDP Needs. Summary of Activities under the Supplementary Appeal" (UNHCR, 2007).

²⁵ *Ibid.*, 33.

²⁶ *Ibid.*, 31.

²⁷ *Ibid.*, 30.

²⁸ United Nations High Commissioner for Refugees, SIR, available at <<http://unchr.org/pages/4a0183436.html>>.

²⁹ United Nations High Commissioner for Refugees, "Practical Guide to the Systematic Use of Standards and Indicators in UNCHR Operations" (UNHCR, 2006), 27.

Given the context where the turn to indicators occurs, it is possible to interpret it as a reaction to the need of rationality and predictability in the UNHCR's involvement with situations involving IDPs. In a context where the need of striking a delicate balance between humanitarianism and sovereignty is imperative, where "hard" norms of international law seem to have limited use, indicators would seem a useful source of justification for UNHCR's involvement with a certain crisis, or to further the agency's involvement. "Hard data" comes in handy for an agency whose domestic presence in an IDP situation needs to be always justified.

That is, though, only part of the story. While the state is still primarily responsible for IDPs, the general use of indicators changes the way the problem of displaced population is dealt with by the national administration. While the UNHCR has no formal authority over the state concerning its national IDP policy, the implementation of indicators as the language of communication between the national state and the international agency provides the latter with a mechanism of, in effect, monitoring domestic policies. This use of indicators has already been noted in other contexts. Human rights indicators, for instance, allow monitoring of compliance and fulfillment of human rights obligations.³⁰ IDP indicators seem to be useful in much the same manner. For example, the Internal Displacement Monitoring Centre keeps rankings of the numbers of displaced people in different states.³¹ It seems unlikely that domestic administrations are unaware of such comparisons.³² For better or for worse, once IDP policy becomes a matter to be expressed in the language of quantitative terms, something changes in the way everyone *thinks* about it.

Indicators in the context of global IDP policy serve, therefore, two different purposes. On the one hand, they are useful to justify international involvement in a situation that is, in principle, recognized as merely domestic. Moreover, indicators are a useful mechanism to influence political outcomes. In this latter sense, indicators seem to work in a way similar to soft law instruments. They are not menaces backed by armed enforcement, but rather seem to change agents' terms of engagement with the problem. Just as soft law norms have proven effective in the IDP context, indicators influence the way domestic governments behave with regards to their IDPs. One could plausibly understand indicators as governance

³⁰ On the audit function of indicators, see AnnJanette Rosga and Margaret Satterthwaite, "The Trust in Indicators: Measuring Human Rights," *Berkeley Journal of International Law* 27 (2009): 279.

³¹ See <<http://www.internal-displacement.org>>.

³² A case in point is the Colombian Acción Social agency, responsible for part of IDP policy in that country. The agency proved concerned enough with rankings, that it adopted at least one official press communiqué entitled: "Colombia is Not Number One in Displacement." It follows: "In connection to international press reports, attributing to the UN information to the effect that Colombia is the country with the largest number of displaced population in the world, the Presidential Agency for Social Action and International Cooperation wishes to state: [...] There is no report by a UN Agency stating that Colombia has the largest displacement in the world [...]." The communiqué then proceeds to compare the amount of IDPs in Colombia, with those in Afghanistan, Iraq and the DRC. See Acción Social, "Colombia no es el Número Uno en Desplazamiento: Acción Social," available at <<http://www.accionsocial.gov.co/contenido/contenido.aspx?conID=5308&catID=127>>; copy on file with the author.

through knowledge and as an exercise of global public authority, in the sense used by von Bogdandy and Goldmann.³³

This angle is certainly important; however, it seems to imply that once the global decision (in this case, the indicator) has reached the domestic setting, domestic authorities and politics have a take-it-or-leave-it choice. Either they follow it, and change their domestic policies accordingly, or they resist it, and follow their home-grown policy in the name of cultural exceptionalism, resistance to neo-liberalism, or other formula.

The goal of this chapter is to complement such an assessment. I suggest, through the case of indicators and IDP policy, that the process of influence is much more subtle: there is a local use of the global indicator, which is then deployed in the domestic setting in order to achieve goals set by the domestic agenda. In that process, the indicator may be transformed, or used for purposes very different from those originally intended. In the case of indicators, there is indeed an effective influence in domestic decisions (and in that sense, an actual exercise of public authority); and yet, that influence is not solely a top-down process, but is rather subject to a subtle process of transformation in the context of use. We shall explore this process in the Colombian context. Before we do that, though, it seems useful to introduce the basic elements of the IDP situation in that country, in order to turn later to the use of indicators therein.

IDPs in Colombia

Armed conflict is the main reason behind internal displacement in Colombia, with armed groups ranging from guerillas, paramilitary groups, and official armed forces, to armies organized by emerald dealers.³⁴ There are documented episodes of internal displacement long before this, during the early 1950s, caused by civil unrest in a period so violent that—standing out even in the violent history of Colombia—it came to be known as *La Violencia*.³⁵ Some contemporary analyses count up to two million displaced people between 1946 and 1966.³⁶ These

³³ See Armin von Bogdandy and Matthias Goldmann, “The Exercise of International Public Authority through National Policy Assessment: The OECD’s PISA Policy as a Paradigm for a New International Standard Instrument,” *International Organizations Law Review* 5 (2008): 241.

³⁴ Procuraduría General de la Nación, *Desplazamiento Forzoso y Reubicación* (1998).

³⁵ La Violencia refers the period of time that began in 1948, with the murder of liberal populist and charismatic leader Jorge Eliecer Gaitán, and ended in 1953, with the coup by General Rojas Pinilla. The murder sparked riots and killings, first in Bogotá and then throughout the country, along political lines: conservative governmental and para-governmental forces against members of the liberal party. Contemporary guerrillas would emerge, in part, from the remains of those original liberal armies. The five years of La Violencia left around 250,000 deaths, in a country whose total population was around nine million at the time. For a useful introduction to La Violencia in English, see Marco Palacio, *Between Legitimacy and Violence: A History of Colombia 1875–2002* (Durham: Duke University Press Books 2006), 135–69. Also useful is David Bushnell, *The Making of Modern Colombia: A Nation in Spite of Itself* (California: University of California Press, 1993), 201–22.

³⁶ Mary Roldán, *A Sangre y Fuego: La Violencia en Antioquia, Colombia, 1946–1953* (Bogotá: Instituto Colombiano de Antropología e Historia, 2003).

episodes, though, were not defined as internal displacement. The contemporary crisis of IDPs started during the late 1990s, and seems to coincide with the moment of highest growth of paramilitary activities and the failure of peace negotiations with the FARC guerrillas.³⁷

To be sure, displacement in Colombia also features an international aspect. The UNHCR reports that there are 77,000 refugees of Colombian origin in the world, and close to 296,000 people in a “similar situation.”³⁸ The IDP situation, though, seems to be more pressing on the domestic political front. By definition, refugees are outside state borders, making them a political liability for the administration in the receiving state.³⁹ In contrast, IDPs have grown to be a sensitive domestic political issue in Colombia.

Internal displacement in Colombia differs from the situation in other countries, where people are massively displaced by a single significant armed action. In Colombia, displacement often occurs individually, as the head of the household is threatened or killed by armed actors in targeted actions. As a consequence, he or she (or his/her survivors) is (or are) forced to move.⁴⁰ The result of this dynamic is that the displaced population increased marginally over the years, and the magnitude of the problem only became apparent after years of this strategy of retail violence. Moreover, this structure led to IDPs being originally considered as homeless individuals in the main cities; that is, as a symptom of the general economic problem of urban poverty, and not as the specific byproduct of a political armed conflict in the rural areas. While data is controversial (a point discussed at length later on), the general perception of the human tragedy suffered by a multitude of Colombians only became a matter of mainstream concern in the late 1990s.

To some degree, the same could be said about the level of awareness in the Colombian government. As early as 1995, the administration recognized its deficiency when dealing with the IDP issue.⁴¹ It lacked a general view of the problem,

³⁷ See Ana María Ibañez, *El Desplazamiento Forzoso en Colombia: Un Camino Sin Retorno Hacia a Pobreza* (Bogotá: Universidad de los Andes, Facultad de Economía, CEDE, Ediciones Uniandes, 2008), 10.

³⁸ SUNCHR, “Refugiados, solicitantes de asilo, desplazados internos, y otras personas bajo el mandato del ACNUR en las Américas,” (SUNCHR, 2008), available at <<http://www.acnur.org/t3/recursos/estadisticas>>.

³⁹ This is quite clear in dealing with Colombia refugees who have fled to Ecuador. According to the UNCHR, there are 19,098 refugees and 82,300 people in “similar situation” in Ecuador, most of whom are of Colombian origin. Ecuadorian President Correa has started to feel political pressure in northern Ecuadorian states, at the border with Colombia, where most of the Colombian refugees are located. These are poverty stricken districts, where refugees come to share poverty with the locals. President Correa has recently called for the “co-responsibility” of the Colombian government with regards to costs. See “Ecuador pide Corresponsabilidad de Colombia Frente a Refugiados,” *Revista Semana*, August 25, 2010.

⁴⁰ See Ibañez, *El Desplazamiento Forzoso en Colombia* (n. 37), 13. It should be noted, though, that as violence increased in the areas of Antioquia, Chocó, and Cesar in 2001, cases of massive displacement became more common. The peak of massive displacement was, according to Ibañez, in 2001–2002.

⁴¹ Conpes document 2804 of 1995. The Conpes (the National Council for Economic and Social Policy) is a legally established entity which serves as a consulting agency for the government in all of its aspects of economic and social policy. It produces several position papers named “Conpes documents,”

and seemed (or at least appeared to be) unaware of its magnitude. At the time, very little government assistance existed for IDPs, and the few aids that were available worked in an isolated and uncoordinated manner. There was neither an articulate public policy nor a coordinated group of institutions to execute it.⁴²

After the official recognition of the unresolved IDP crisis, the government created the National Program for Integral Attention to the Displaced Population due to Violence (the "Program").⁴³ Not long after this initial response to internal displacement, the government recognized that the Program had not been successful, as it had shown "inter-institutional management and coordination difficulties, as well as financial and information deficiencies."⁴⁴ In consequence, the government created the National System for the Integral Attention for Displaced Population due to Violence. The new system attempted to correct the mistakes made by earlier attempts to make a blanket approach to the IDP crisis by rearranging some of the existing policies. Most importantly, it also called for a meeting of the National Council for the Displaced Population, which had existed for two years but had failed to meet.⁴⁵

Notwithstanding the government's attempts at tackling the IDP problem, its efforts were met with legislation that sought to institutionalize all coordinated approaches.⁴⁶ In July 1997, Congress enacted Law 387 of 1997 (the "Internal Displacement Attention Act"), which served as the legal framework for the integral aid that should be offered to IDPs, and was the first legal recognition of the rights they were entitled to.⁴⁷ Arguably, the Act had some impact on the drafting of Deng's Guiding Principles on Internal Displacement,⁴⁸ adopted only a few months later, in February 1998.⁴⁹ The Act was significant, in that it finally established an actual policy that gave visibility to IDPs as a distinct group of the population, different from other victims of other "emergencies," that was entitled to attention to its special needs.⁵⁰ However, its application was limited, and the government

which embody the decisions and recommendations taken by the national government regarding the areas of its jurisdiction.

⁴² Cesar Rodríguez and Diana Rodríguez, "El Contexto: El Desplazamiento Forzado y la Intervención de la Corte Constitucional (1995–2009)," in Cesar Rodríguez (ed.), *Más allá del desplazamiento: políticas, derechos y superación del desplazamiento forzado en Colombia* (Bogotá: Ediciones Uniandes, 2009), 19.

⁴³ Conpes document 2804 of 1995.

⁴⁴ Conpes document 2924 of 1997.

⁴⁵ Ibid.

⁴⁶ C. Rodríguez and D. Rodríguez, "El Contexto: El Desplazamiento Forzado y la Intervención de la Corte Constitucional" (n. 42), 20.

⁴⁷ Law 387 of 1997 (July 18). Published in the Official Diary no. 43,091, July 24, 1997.

⁴⁸ *Guiding Principles on Internal Displacement*, E/CN.4/1998/53/Add.2, February 11, 1998.

⁴⁹ This argument has been suggested before in Rodríguez and Rodríguez, "El Contexto: El Desplazamiento Forzado y la Intervención de la Corte Constitucional" (n. 42), 21.

⁵⁰ Manuel José Cepeda, "The Constitutional Protection of IDPs in Colombia," in Rodolfo Arango (ed.), *Judicial Protection of Internally Displaced Persons: The Colombian Experience* (Washington: The Brookings Institution—University of Bern Project on Internal Displacement, 2009), 7.

acknowledged that serious modifications were needed.⁵¹ In consequence, Decree 2569 was issued, attempting to further regulate the Act. The Decree undertook several measures, most notably creating a Single Registry of Displaced Population, that attempted to register information on the number and condition of the displaced population of the entire country. Additionally, it created municipal, district, and state committees to monitor and aid the displaced population as provided for in Law 387.

In spite of the amount of legislation, regulation, and executive measures to deal with IDPs, the late 1990s were critical for the displaced population. Specifically, they were not enjoying their fundamental rights provided by the Constitution. This led to the landmark decision by the Colombian Constitutional Court that changed the IDP debate, and brought indicators into the spotlight. To understand the context of the decision, it is first useful to briefly explore how the debate on IDPs ended up in the hands of the Constitutional Court, and the role of indicators in that context. Under Colombian constitutional law, there is an exclusive constitutional jurisdiction, which is headed by the Constitutional Court, a high court with the final word on constitutional matters. The constitutional jurisdiction decides *acciones de tutela*, a legal action that seeks an immediate judicial injunction against the breach of fundamental constitutional rights (such as the right to life, the right to privacy, due process, etc.). Such action may be filed by any person who believes that his or her fundamental rights have been infringed, and has an expedited process whereby judges must render a decision on the matter within ten days of its filing. A decision rendered on an *acción de tutela* has an appeal procedure, and may be eventually selected for *certiorari* by the Constitutional Court.

Given the severe conditions faced by IDPs, *tutelas* became a preferred mechanism of judicial protection. By the time that the Constitutional Court rendered its influential decision on IDPs in 2004, it had already ruled 17 cases dealing with the rights of the IDPs: (i) in three cases to protect them against discrimination; (ii) in five cases to protect their life and personal integrity; (iii) in six cases to guarantee the effective access to health; (iv) in five cases to protect the right to the minimum income standard; (v) in two cases to protect the right to housing; (vi) in one case to protect the right of free transit; (vii) in nine cases to protect the right to an education; (viii) in three cases to protect the rights of children; (ix) in two cases to protect the right to choose one's own domicile; (x) in three cases to protect the right to work; (xi) in three cases to protect the right of access to humanitarian emergency aid; (xii) in seven cases to avoid the prerequisite of registry in the IDP database as a condition for government aid.⁵²

⁵¹ The weaknesses of the existing regulation at the time were expressed through Conpes document 3057 of 1999, which identified as a main failure the lack of a single methodology to gather data and indicators to assess relevant information and prevent further displacement.

⁵² Rodríguez and Rodríguez, "El Contexto: El Desplazamiento Forzado y la Intervención de la Corte Constitucional" (n. 42), 25.

By 2004, the amount of *acciones de tutela* being filed by displaced persons reflected a general and structural problem, which is why the Constitutional Court decided to review them. When case file no. T-653010 reached its docket, involving a petition for protection regarding a displaced person, the Court decided to join 108 cases, which were filed by 1,150 families (roughly 4,000 claimants), all of which were under a similar situation of constitutional breach because of their displaced condition.⁵³ The result was Decision T-025 of 2004, which declared that internal displacement had become an *unconstitutional state of affairs*⁵⁴ and ordered many public institutions to undertake several changes in their policy, in order to give adequate coverage to IDPs, protecting their constitutional rights.

IDP indicators enter the domestic scene

The decision taken by the Court is notable in at least four senses.⁵⁵ First, its sheer ambition is remarkable. This is a decision of enormous implications, considering the number of people whose rights are protected, the number of public agencies called into action by the Court, and the amount of resources (economic and otherwise) required to comply with the Court's orders. Second, the Court's decision neatly falls on the global governance nature of IDP policy. While clearly dealing with a matter of domestic law and policy, and addressed mainly to domestic bureaucracies, the decision drew heavily from international law; mainly, Deng's Guiding Principles on Internal Displacement. For the Court, the Principles were "pertinent for the interpretation" of rights provided in the domestic constitution.⁵⁶ Thus, the Court's methodology was to establish a series of IDP rights under domestic law that were threatened by the unconstitutional state of affairs (say,

⁵³ The decision to join these files is given by a writ dated November 10, 2003. I translate as "writ" the term Spanish term "Auto," which refers to a judicial decision that decides matters that are not of substantive importance for the conflict. Under Colombian procedural law, there are no specific writs ordered by the Court, but only a general term which serves to refer to every order of a Judge which is not the ruling of a case.

⁵⁴ Decision T-025 of 2004. An unconstitutional state of affairs is a legal doctrine under Colombian constitutional law whereby it is declared that a violation of fundamental rights is systematic, widespread, and due to structural causes, thus warranting a judicial intervention on general policy. Perhaps the closest notion would be that of structural injunction applied, *inter alia*, in US and South African constitutional law. A structural injunction is a "formal medium through which the judiciary seeks to reorganize ongoing bureaucratic organizations so as to bring them into conformity with the Constitution" (Owen Fiss, "The Allure of Individualism," *Iowa Law Review* 78 (1993): 965). For an introductory piece on the US side, see Owen Fiss, *The Civil Rights Injunction* (1978), 7; on the South African approach, see Danielle Hirsch, "A Defense of Structural Injunctive Remedies in South African Law" *Oregon Review of International Law* 9 (2007): 10–18. In the Colombian case, the "unconstitutional state of affairs" was declared as the Court found that the IDP situation constituted a repeated and constant violation of fundamental rights, affecting a multitude of persons, due to problems of a structural nature and requiring the intervention of several state authorities for its resolution (see Cepeda, "The Constitutional Protection of IDPs in Colombia" (n. 50), 18).

⁵⁵ For arguments one, three, and four, see Cesar Rodríguez and Diana Rodríguez, *Las Cortes y el Cambio Social Estructural: Los Efectos del Constitucionalismo Progresista* (Bogotá: DeJusticia, 2010), 15.

⁵⁶ Decision T-025 of 2004, Section 2.1.3.

the “right to life in dignified conditions”⁵⁷) and then find the Guiding Principles that were pertinent for interpreting such domestic rights (in our example, Principles 1, 8, 10, and 13⁵⁸). In that way, the Court skillfully articulates the global nature of its undertaking. Third, the Constitutional Court did not abandon the IDP problem merely by rendering its far-reaching decision. It decided to remain involved in the matter⁵⁹ and traced its effects and compliance strictly; quite an unusual move for this kind of rulings.⁶⁰ And, finally, beyond simply deciding that certain rights were violated, the Court felt it necessary to hold that “No specific goals or indicators have been established, which can allow for a verification of whether the purposes of the policy have been fulfilled or not,”⁶¹ thus ordering that IDPs indicators be created.⁶²

In 1994, Francis Deng, the Special Representative of the UN Secretary-General for IDPs, had already pointed out that the Colombian government had no indicators at all regarding internal displacement, and that the gravity of the problem had been established merely by word of mouth.⁶³ He had also noted that there was no effort to create a methodological approach towards the creation of indicators on internal displacement.⁶⁴ In the following years, three different attempts to create information systems on IDPs were made. The first of these systems, devised in 1997, was the RUT System of the Colombian Episcopal Conference. Its methodology to recover information is built around churches and religious establishments, where IDPs arrive and share their own information with the religious leader of the community. Although there is no information-gathering on persons who fail to arrive at these Catholic establishments, there seems to be wide coverage on the whole territory, as well as general access to the displaced population.⁶⁵ A second

⁵⁷ Decision T-025 of 2004, Section 5.2.1.

⁵⁸ *Ibid.*

⁵⁹ Article 27 of Decree 2551 of 1991, which regulates the exercise of jurisdiction of the Constitutional Court, allows it to remain seized of a matter until the breach has ceased to exist.

⁶⁰ It is worth noting that the Court went beyond using writs as mere procedural orders, and implemented what it called “Autos de Seguimiento” (more or less: “Follow-up Writs”) under which the Court supervised compliance with the main decision. It is reasonable to say that, without these follow-up writs, the Court’s order of indicators would have remained a dead letter.

⁶¹ Decision T-025 of 2004, Section 6.3.1.1(ii). In the same sense, according to Section 6.3.1.3(ii) “[...] there do not exist systems to evaluate the policy. The policy does not include a system designed to detect mistakes or obstacles in its design and implementation, needless to say one that allows an adequate and timely correction of such failures. There are no systems or indicators for the verification, follow-up and evaluation of results, either at the national or territorial levels” (footnotes omitted, RU).

⁶² Specifically, the Court found that part of the problem was that there were deep institutional flaws in IDP policy, one of which was lack of indicators. Therefore, the Court ordered the relevant agencies “to adopt, within the three months following the communication of this judgment, a program of action, with a precise schedule, aimed at correcting the flaws in institutional capacity, at least with regard to the ones indicated in the reports that were incorporated to the present process and summarized in Section 6 and Annex 5 of this Judgment” (Decision T-025/94, Order 4).

⁶³ United Nations Commission on Human Rights, Internally Displaced Persons: Report of the Representative of the Secretary General, Francis Deng, submitted pursuant to CHR resolution 1993/95 Addendum 1: Profiles in Displacement: Colombia, 1994, UN Doc. E/CN.4/1995/50/Add.1, para. 10.

⁶⁴ *Ibid.*, para. 15.

⁶⁵ While the government does not keep religious data, recent media polls have shown that the country is close to 81 percent Catholic. The Catholic Church is often the main institutional presence

effort in compiling data was attempted by the International Committee for the Red Cross (ICRC), though restricted only to its humanitarian aid operations. The manner by which the ICRC collects data is reliable, including domiciliary visits and personalized interviews. Its scope, however, is limited, as it may only collect information on the displaced persons under the ICRC's attention, which fails to cover the total of IDPs in the country. The final data producer is the government, through its Unique Registry System of the Social Solidarity Network, created by the Law 387. This system only stores the information of displaced persons who have approached official entities to declare their situation.

Enter the 2004 decision. The Court based its reasoning on the idea that the mere re-statement of rights was insufficient to solve the unconstitutional state of affairs found in the IDP situation in Colombia. It therefore adopted the "effective enjoyment of rights" as the relevant criterion for evaluation—a specific results-based threshold to be achieved by IDP policy in Colombia. This focus in results (and not in processes) was reflected in the Court's approach to indicators. It soon became evident to the Court that effective implementation of its orders could not be measured in terms of bureaucratic processes (that is, meeting after meeting of the relevant actors with no results in sight), but should rather be measured in terms of results. However, despite the several attempts referred to before, by the time the decision was rendered there were very few results indicators in national or international debates. In fact, the best evidence of this is the relative absence of reliable or independent indicators in the text of the decision itself.⁶⁶

The bottom line was that the results-based "effective enjoyment of rights" approach required some sort of indicators, which were wholly absent from the Colombian landscape. Therefore, the Court ordered the development of such "results indicators," as opposed to "process indicators," making the issue a central tenet of the follow-up process. At least seven of the writs following the decision were concerned with indicators.⁶⁷ In the first couple of years after the decision, it became clear that the institutions were not developing the indicators wanted by the Court.⁶⁸ Therefore, in 2006, the Court established a fast-track procedure for the creation of indicators, and called upon the UNHCR, the Colombian Ombudsman, and other institutions to present proposals for indicators that would do the trick.⁶⁹

in areas of conflict. Information collected by these Catholic institutions were, therefore, the first clear source of hard-date concerning IDPs. See "¿Qué tan Católicos Somos" in *El Tiempo*, April 8, 2001. On the Church's role in the armed conflict, see generally: R. Arias and F. González, "Búsqueda de la Paz y Defensa del 'Orden Cristiano': El Episcopado ante los Grandes Debates de Colombia (1998–2005)", in Francisco Leal Buitrago (ed.), *En la Encrucijada: Colombia en el Siglo XXI* (Bogotá: Norma, 2006).

⁶⁶ Decision T-025 of 2004, Section 6.3.1.1(ii) and 6.3.1.3(ii). The very first indicator referred to by the Constitutional Court related to the level of nourishment of displaced populations (n. 86). The Court further relied on governmental statistics and indicators both to determine the amount of displaced persons in Colombia, by using a report from the governmental agency Red de Solidaridad Social (n. 93), and to determine the amount of resources allocated to IDPs, by relying on a report sent by the government itself (n. 115). Furthermore, some indicators, such as the level of access to higher education (at 45) or the "level of fulfillment of basic needs" (at 61) have no referred source.

⁶⁷ Writ 337 of 2006, Writ 027 of 2007, Writ 109 of 2007, Writ 116 of 2008, and Writ 011 of 2009.

⁶⁸ Writ 218 of 2006.

⁶⁹ Writ 337 of 2006.

Over 500 indicators were proposed in this process.⁷⁰ On March 1, 2007, the Court held a public hearing to discuss the proposals.⁷¹ On that date, the government also presented what has been called a “serious”⁷² indicators proposal to the Court. Months later, the Court decided on the proposal, holding that some of the proposed indicators should be rejected, as they were not “adequate, relevant and sufficient” to measure the effective enjoyment of rights, as defined by the Court.⁷³ Other indicators, in turn, did follow the Court’s Guidelines, but had gaps that had to be filled.⁷⁴ And, finally, some indicators could be adopted as they were.⁷⁵ The Court asked the government to submit better indicators to replace those that had been rejected. The Court then decided on this new submission,⁷⁶ and went over the process once more,⁷⁷ thus settling on the final version of the basic toolkit of indicators that would be available to evaluate whether (and when) the unconstitutional state of affairs was overcome.

With the decision, several Colombian institutions were obliged to produce periodical reports with their respective indicators, and send them to the Court. The Tribunal, though, was critical of the work done by governmental agencies in this regard. Particularly, in Writ 218 of 2006, it was critical of the impossibility of determining any improvement in the general situation of IDPs due to the lack of relevant indicators. In contrast, government officials had submitted huge amounts of irrelevant information whilst, according to the Court, the most eloquent measure of compliance was still missing: statistical data that indicated the fulfillment of IDPs’ rights.⁷⁸

At any rate, after six years of follow-up writs, and the mobilization over indicators of several governmental agencies, activists, and academics, the Court ended up with a complex set of indicators to measure 20 IDP rights,⁷⁹ each of which measured from three different perspectives: (a) “effective enjoyment of rights” indicators (which is self-explanatory); (b) “complementary” indicators (which measure the general context where the effective enjoyment is measured); and, finally, (c) sector indicators, measuring the advances of each the governmental

⁷⁰ Clara Elena Reales, “Design and Implementation of the Orders Issued in Decision T-025 of 2004: An Assessment of the Process,” in Arango (ed.), *Judicial Protection of Internally Displaced Persons*, (The Brookings Institute—University of Bern Project on International Displacement 2009), 80.

⁷¹ See Writ 027 of 2007.

⁷² Reales, “Design and Implementation of the Orders Issued in Decision T-025 of 2004” (n. 70), 80.

⁷³ Writ 109 of 2007, paras 57–71.

⁷⁴ Writ 109 of 2007, paras 28–31.

⁷⁵ Writ 109 of 2007, Section 81.3.

⁷⁶ Writ 337 of 2007.

⁷⁷ Writ 116 of 2008.

⁷⁸ Writ 218 of 2006, 3–5.

⁷⁹ The rights measured by indicators are: (1) right to life; (2) personal integrity; (3) freedom; (4) housing; (5) health; (6) education; (7) food; (8) “income creation”; (9) identity; (10) social and economic stability; (11) prevention of further displacement; (12) return; (13) subsistence; (14) family reunification; (15) personal security; (16) participation; (17) truth, justice, and reparation; (18) specific indicator: boys, girls, and teenage IDPs; (18) specific indicator: ethnic and different culture IDPs; and (19) specific indicator: gender IDPs. Source: Cesar Rodriguez “Más Allá Del Desplazamiento, O Cómo Superar Un Estado De Cosas Inconstitucional” in Rodriguez (ed.), *Más Allá del Desplazamiento* (n. 42), 460–73.

agency involved.⁸⁰ Thus, for example, the right to life would be measured by (a) an effective enjoyment of rights indicator (“members of the displaced family remain alive”); (b) complementary indicators (for example, the “number of IDPs killed for causes directly related to displacement as a percentage of the general IDP population”); and (c) several sector indicators (for example, the “number of IDP leaders who are under special police protection as a percentage of the total of IDP leaders”). The idea, then, is that governmental agencies are to provide data to feed these indicators, and will be measured according to them. Only by observing the landscape drawn by these quantitative instruments, will we ever know whether the unconstitutional state of affairs has been solved.

IDP indicators as a technology of global governance

The Court’s order sparked a number of processes and interactions that are revealing of the complex political, institutional, and legal effects implicit in the turn to indicators. This section explores such processes, focusing on (a) domestic IDP indicators as input for global decisions; (b) the role of the news media in the turn to indicators as technologies of global governance; (c) IDP indicators as social mobilization; and, finally, (d) IDP indicators and their role in the interaction among separate branches of power. From such areas of exploration, some conclusions concerning indicators as tools for global governance will be drawn. These conclusions will be presented in the last section of this chapter.

Domestic IDP indicators as input of global decision-making

As we have seen, the central turn to indicators at the UNHCR occurred around 2005, with its cluster approach. It is interesting that, since then, domestically produced IDP indicators have had an impact on international monitoring. In the fleeting reference made to IDPs in the Report of the UNHCR on the situation of human rights in Colombia in 2008,⁸¹ the Commissioner relied exclusively on the figures presented both by an independent NGO (CODHES) and the Colombian government itself (Social Action).⁸² The report presented by the Norwegian Refugee Council’s Internal Displacement Monitoring Centre (IDMC) for the consideration of the Committee for the Elimination of Racial Discrimination at its 75th session⁸³ made the disparity in indicators evident, and relied heavily on the requests of the Constitutional Court for the compliance of Decision T-025,⁸⁴ which, in turn, were framed almost exclusively on the demands for the production

⁸⁰ *Id.*

⁸¹ UN Doc. A/HRC/10/032 (draft of December 31, 2008).

⁸² *Ibid.*, 21.

⁸³ Available at <<http://www.internaldisplacement.org/8025708F004CE90B/%28httpCountries%29/CB6FF99A94F70AED802570A7004CEC41?OpenDocument>>.

⁸⁴ Report of the Displacement Monitoring Centre (IDMC) for the consideration of the Committee for the Elimination of Racial Discrimination at its 75th session, para. 1.

of indicators. This too was the case for the report presented to the Human Rights Committee.⁸⁵

It makes sense, then, that the Court felt obliged to order the production of indicators to aid in the determination of compliance on behalf of the government. As a result, CODHES endorsed another independent commission to collect direct evidence on displaced populations: the Monitoring Commission of Forced Displacement Public Policy, composed of renowned academics, jurists, and economists. The Commission produced its first report in 2006, and has produced 15 reports since then with its own methodological and statistical approach, as well as independent data gathering.

In the international sphere, the most readily available database on indicators of forced displacement is the already cited IDMC of the Norwegian Refugee Council, which relies almost exclusively on the indicators provided by CODHES and its monitoring Commission, both on its website and on the reports submitted to international organizations such as the United Nations.⁸⁶ Furthermore, the UNHCR refers to both CODHES and governmental figures in its annual report, and takes no particular position on disputed figures.⁸⁷ It seems clear that the UN and its subsidiary organs, at least, are the consumers of locally produced indicators.

The local character of these indicators should not be exaggerated. The whole process of the follow-up writs implied an interaction of domestic agencies with the UNHCR. Moreover, domestic NGOs were involved, whose funding and network may very well be global in origin and reach. However, it is clear that the origin of the IDP indicators was not a global governance institution, which then imposed it or transplanted it into the domestic setting. In that sense, even if the local process was not clinically isolated from global influences, it may be still meaningfully understood as local. This dynamics problematizes the governed–governors–public framework put forward by Davis, Kingsbury, and Merry.⁸⁸ It seems awkward to think of the UNHCR as the governor in this context. Rather, it was the Court's emphasis on the effective enjoyment of rights and the subsequent "results" rhetoric that brought emphasis to the quantitative tools of governance. The UNHCR, of course, followed suit. The role of governmental agencies seems to be an ill fit here as well, as they seem to be both governors and governed at the same time. The Court, in turn, seems to be a governor, despite being part of a state that would be one of the "governed." Moreover, the indicators were created by a

⁸⁵ Submission from the Internal Displacement Monitoring Centre (IDMC) of the Norwegian Refugee Council (NRC) for consideration in the formulation of the List of Issues by the Human Rights Committee's Task Force during the 97th Session (October 12–30, 2009).

⁸⁶ See <<http://www.internal-displacement.org>>.

⁸⁷ Annual Report of the United Nations High Commissioner for Human Rights and Reports of the Office of the High Commissioner and of the Secretary General—Report of the United Nations High Commissioner for Human Rights on the Situation of Human Rights in Colombia, UN Doc. No. A/HRC/10/032, para. 21.

⁸⁸ Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, "Indicators as a Technology of Global Governance," *Law and Society Review* 36:1 (2012): 71–104, esp. 78–90.

process of cooperation between NGOs, the UNHCR, governmental agencies, and the Court itself. Who is the governed here? Who is the governor?

The role of NGOs and the media seem also to reveal the limits of the model. Activist mobilization played a central role in defining the specific traits of technology of global governance. If NGOs are active in defining the relevant indicators, can they still be considered part of the "public"? The same can be said about the media. Media coverage plays a crucial role in the emergence of an indicator as a veritable technology of global governance, beyond it being a mere statistical piece of information. If that is the case, can the media still be considered as part of the "public"? Does it not have a more direct stake in the development of this particular exercise of global power? And yet, should we then understand the media as part of the governing structure? Is it possible to draw a line between the public and the governor if the latter is wholly dependent on the former? Part of the answer to these questions may come from taking a closer look at the role of the media in the Colombian IDP case.

IDP indicators and the news media

The authority of IDP indicators is intimately linked with the media.⁸⁹ Reference to IDP indicators has been a key point in public media coverage at least since the late 1990s, after major legislative changes concerning public displacement took place. It is noteworthy that the manner in which the media refers to IDP indicators plays an important role in their function as technologies of global governance. There seem to be at least three different approaches to indicators in the media: (1) indicators with no stated source (presumably taken from other media); (2) indicators with a stated source (regardless of the source itself, e.g., the government or independent NGOs); and (3) indicators with a misleading source (which prove to be the most interesting case of creative use, as will be seen further on).

As early as 2000, newspapers were adopting the first approach to IDP indicators: high numbers with no reference to sources. At the time, newspapers were reporting indicators as high as 1.5 million IDPs in Colombia.⁹⁰ The second use of indicators is also common at this point, using, for example, indicators produced by some of the actors discussed earlier.⁹¹ It is important to note, however, that prior to the

⁸⁹ The foregoing section refers almost exclusively to references made in published media in Colombia. The pool of available media includes the seven newspapers with highest circulation in the country for the last three years, as well as the two most-read news sources for the past ten years (in magazine and newspaper format). References made to these will include the publication name and date, as most of the sources were taken from digital format. Where reference to a specific published material is available, it is expressly mentioned. All quotations taken from these sources have been freely translated by the author into English.

⁹⁰ "Mil Millones de Pesos de la Red Para Desplazados del Sucre," *El Tiempo*, January 1, 2000.

⁹¹ For example, the ecclesiastical production of indicators by the Pastoral Social of the Catholic Church. See "Desplazados no Saldrán del Pastoral," *El Tiempo*, July 27, 2000. This approach is not missing entirely from latter dates, where the reporting style on indicators had changed. See, e.g., "Desplazados, una bomba de tiempo," *El Tiempo*, December 22, 2004.

Constitutional Court's intervention in 2004, production and use of indicators were of lesser importance and prominence in the public sphere.

By 2004, the third approach to indicators in the media took an important turn. As we have seen, the production of statistical data on IDP was a major concern in the Court's decision in 2004, and throughout the process that followed. Once referred to in the text of the ruling, these numbers acquired an authority for the media that did not exist beforehand. All of a sudden, figures that used to be read with a grain of salt, became "what the Court said," and thus, acquired a higher sense of reliability by magically changing their producer.⁹² It is also around this time that the media started to report the difference in indicators depending on their source, quite literally referring to the phenomenon as "quarrels over indicators,"⁹³ probably due to the Court's admonishment on this issue. Global decision-makers, such as the UNHCR, became empowered by this move, and played a more important part in the media during and after 2004, considering the increased spotlight placed by the Court on the production of indicators, even on an international level.⁹⁴

As already seen, the Court's decision in 2004 radically changed the debate on IDPs and indicators. Media reference to the topic took two major paths after 2004: much like the Court itself, newspapers were either chastising public institutions due to their lack of compliance with decision T-025,⁹⁵ or there was a general admiration for the Court's action, seen as a bold step towards social justice.⁹⁶ It was not until the mid and late 2000s that media reports on IDP indicators were spiced up by the addition of international reports on the same topic. Not only the UNHCR, but other international institutions, such as the Norwegian Refugee Council through its IDMC, started to do some heavy reporting on the IDP situation. During this time, the way in which the media wrote about internal displacement and the sources used for their indicators were much more sophisticated. A general tendency to discriminate sources to attempt to draw an objective picture emerged.⁹⁷ Furthermore, reliance on other indicator producers became more typical to include, for example, third-party NGOs.⁹⁸ In these cases, claims on the number of IDPs in Colombia were generally of high proportions and made by press statements and not through published materials.⁹⁹

⁹² See, e.g., "54 Días para dar respuesta a desterrados," *El Tiempo*, February 7, 2004; "No Más Pañitos de Agua Tibia," *Revista Semana*, February 9, 2004; "Se agota el plazo," *Revista Semana*, March 12, 2004.

⁹³ "Agarrón por las cifras," *El Tiempo*, September 3, 2004.

⁹⁴ "Acnur pide replantear política de desplazados," *El Tiempo*, December 16, 2004.

⁹⁵ "No hay política de desplazados," *El Tiempo*, May 12, 2004; "Crisis en los desplazados del meta," *El Tiempo*, October 12, 2004.

⁹⁶ See "Desplazados," *Revista Semana*, July 30, 2004.

⁹⁷ "Comunidades afro de Tumaco urgen atención del gobierno," *Revista Hechos del Callejón*, March 18, 2008. It is noteworthy that this is the first reported reference to the Norwegian Refugee Council's use of indicators on IDPs.

⁹⁸ In this particular case, an Ecuadorian religious NGO (the "Jesuit service for refugees and migrants") is quoted by: "ONG asegura que hay más de 600 mil desplazados colombianos," *El Espectador*, March 14, 2008.

⁹⁹ *Idem*. According to *El País*, the Ecuadorian office of the UNHCR noted that there were 3.5 million internally displaced persons in Colombia by 2008 through a public statement.

The most cautious approach was taken by the ICRC, perhaps in accordance with their mandate in the country. When interviewed, the spokesman for the ICRC described merely a range of indicators and was careful to explain that his organization did not have "competence to refer to the amount of internally displaced persons" but merely to the amount of people to which they had provided humanitarian aid.¹⁰⁰ Nonetheless, the Colombian media attempted to rely on the ICRC as a producer of indicators, even on the total amount of IDPs, going as far as to refer to their statements as "reports."¹⁰¹

At this point, the different sources of indicators and their political implications had already been noted by the national and international press. However, reports from international monitoring bodies started to emerge, quickly shifting the manner in which these indicators were seen in the Colombian media. During 2008, at least five different major Colombian newspapers began relying on indicators produced by international entities. *CMI*, a Colombian news network, reported that the UNHCR had presented a report in Geneva which stated that there were four million IDPs in Colombia.¹⁰² *El Universal*, a regional newspaper of importance along the Caribbean coast, referred to a report of the IDMC which also stated that there were four million IDPs in the country,¹⁰³ using the terms "IDMC" and "UNHCR" interchangeably. *El Espectador*, which is largely seen by Colombian journalist circles as being an independent newspaper, referred to the same report produced by the IDMC but this time "under the auspices" of the UNHCR.¹⁰⁴ *El Tiempo*, by all accounts the most influential newspaper of the country, referred to (presumably) the same report, although this time as authored by the Norwegian Refugee Council and rendered "in the presence of the UNHCR" (emphasis added), again with the same four million figure.¹⁰⁵ *El Heraldo*, another influential Caribbean newspaper, wrote of the report presented by the IDMC with the same figure,¹⁰⁶ and *El País*, the leading newspaper in the western side of the country and the Pacific coast, reported the same figure, although attributing the report both to the UNHCR and the IDMC.¹⁰⁷

All of these publications were referring to the only relevant report rendered at the time: the IDMC Activity Report of 2007, published in April of 2008, which in turn referred to a country-specific report entitled "*Resisting Displacement by Combatants and Developers: Humanitarian Zones in North West Colombia*." Although the report did have a fleeting reference to indicators by stating that "almost four million

¹⁰⁰ "Desplazarse no es solo huir, es perderlo todo," *British Broadcasting Company (BBC)*, April 2, 2008.

¹⁰¹ "Aumento del desplazamiento en diferentes regiones del sur occidente colombiano," *El Tiempo*, April 2, 2008.

¹⁰² "Sigue aumentando en Colombia número de desplazados según ONG," *CMI*, April 16, 2008.

¹⁰³ "Situación de desplazados empeoró en 2007," *El Universal*, April 16, 2008.

¹⁰⁴ "Conflictos internos dejaron 26 millones de personas desplazadas," *El Espectador*, April 17, 2008.

¹⁰⁵ "Hay 26 millones de personas desplazadas en el mundo," *El Tiempo*, April 17, 2008.

¹⁰⁶ "Desplazamiento se agravó," *El Heraldo*, April 17, 2008.

¹⁰⁷ "Colombia, el segundo país con más desplazados. Junto a Iraq y Sudán lidera la lista de los países con alto número de desplazados," *El País*, April 18, 2008.

Colombians have been displaced,” its source had quite a different take on the matter. The IDMC report is exhaustive on its use of indicators, not only describing in full detail the production of indicators in Colombia and the general issues with their use as warned by the Constitutional Court, but being cautious on its very use of indicators. Its general overview of IDP indicators in Colombia warns that “there is a huge discrepancy between official and non-official IDP numbers reflecting different approaches to durable solutions, the nature of the conflict and to what constitutes the end of displacement.”¹⁰⁸ It then summarizes the divergence of indicators in Colombia: “The Colombian NGO CODHES has recorded a total of 3.7 million internally displaced in Colombia since 1985. The government’s current estimate is much lower, at 1.75 million, largely due to the fact that it started registering IDPs systematically only in 2000. CODHES registered about 1.8 million IDPs between 1985 and 1999, whereas the government in the same period registered only 79,000. However, from 2000, the year the government improved its registration procedures, there is much less discrepancy. CODHES registered around 1.8 million between 2000 and October 2005, against 1.6 million recorded by the government.”¹⁰⁹

The same discrepancy on indicators reported by international bodies continued in the late 2008 as quoted by the Colombian press. According to *Semana*, the leading magazine on political affairs, the UNHCR had “reported over 3 million internally displaced persons.”¹¹⁰ This same figure was reported by *El Espectador* when describing the constant quarrel between the government and independent organizations as to the amount of displaced persons in the country,¹¹¹ and furthermore by *El País*,¹¹² the online media *Terra*,¹¹³ and the international press agency *Reuters*.¹¹⁴ It remains unclear why the same newspapers that had published similar stories that same year, reported a different indicator a few months later, though the various sources referred to statements made by UNHCR in Europe. In this context, it seems of importance to note that only two UNHCR reports were produced at the time: the *Annual Report of the United Nations High Commissioner for Human Rights and Reports of the Office of the High Commissioner and of the Secretary General—Report of the United Nations High Commissioner for Human Rights on the situation of human rights in Colombia*,¹¹⁵ still in its draft form, and the *UNHCR Global Report 2007—Colombia Situation*,¹¹⁶ which is part of the

¹⁰⁸ Internal Displacement Monitoring Centre, “Colombia: Resisting Displacement by Combatants and Developers: Humanitarian Zones in North-west Colombia. A Profile of the Internal Displacement Situation,” November 5, 2007, 99.

¹⁰⁹ *Ibid.*

¹¹⁰ “La hecatombe de las cifras,” *Revista Semana*, June 21, 2008.

¹¹¹ “Desplazamiento no se puede medir,” *El Espectador*, June 18, 2008.

¹¹² “Acnur aclara que Colombia es el segundo país con más desplazados en el mundo,” *El País*, June 17, 2008.

¹¹³ “ACNUR dice que refugiados y desplazados internos alcanzaron cifra récord 2007,” *Terra*, June 15, 2008.

¹¹⁴ Reuters, “Colombia reporta más de 2.5 millones de desplazados por conflicto,” *Reuters*, June 13, 2008.

¹¹⁵ UN Doc. No. A/HRC/10/032.

¹¹⁶ Available at <<http://www.unhcr.org/484923382.html>>.

UNHCR Fundraising Reports. The former had but a fleeting reference to IDPs, which was introduced in the following manner: "All the figures show a continuous growth in the numbers of IDPs, with a disproportionate effect on communities located in regions where the armed conflict is most intense."¹¹⁷ And, even though no figures were included in the body of the text, such statement was based on the following sources: "Between January and June 2008, the NGO Consultancy on Human Rights and Displacement (CODHES) recorded 270,675 new IDPs, 41 percent higher than during the same period in 2007. According to Acción Social, 249,816 IDPs were registered between January and October 2008."¹¹⁸

On the other hand, the 2007 Global Report included a chart with the relevant indicators on refugees, asylum-seekers, and IDPs, although it actually left blank any IDP indicators with the following caveat-footnote: "According to the Constitutional Court of Colombia, there is a discrepancy between the real number of internally displaced people and the number given by the national registration system. The Court cites the Director of the Agencia Presidencial para la Acción Social y la Cooperación Internacional who acknowledged that the number of IDPs in Colombia is close to *three million* (Order of Compliance 218, dated August 11, 2006, related to the landmark Judgement T-025)"¹¹⁹ (emphasis added). This trend continued in 2009, albeit not with the same thrust as it had in previous years. *El Tiempo* reported that according to the UNHCR, there were 3.3 million IDPs in the country,¹²⁰ while *Semana* reported roughly the same number,¹²¹ both relying on the UNHCR Global Trends 2008 report,¹²² although the report avoided entirely any figure on the number of IDPs and Colombia, and such indicator is nowhere to be found. In the latter years after the rise in international reports concerning the IDP situation in Colombia, a more sophisticated attempt at describing the different figures from different sources was made by the Colombian media. Significantly, both *El Tiempo*¹²³ and *Semana*¹²⁴ made an attempt to more accurately explain the divergence in indicators.

What kind of indicator is interesting for the media? It seems that the media is interested in the indicators that shame the government, either because they show a heartless administration doing nothing for victims of the conflict, an administration being shamed by an international actor, or an inept administration that is unable to have reliable figures on the victims' plight. It is interesting to note, though, that

¹¹⁷ UN Doc. No. A/HRC/10/032, 21.

¹¹⁸ Ibid.

¹¹⁹ UNHCR Global Report 2007—Colombia Situation, 452.

¹²⁰ "Desplazados en Colombia son más de 3.37 millones, dice agencia de refugiados de la ONU," *El Tiempo*, June 16, 2009.

¹²¹ "Colombia, entre los que más desplazados tiene en el mundo," *Revista Semana*, June 6, 2009.

¹²² UNHCR, 2008 Global Trends: Refugees, Asylum-seekers, Returnees, Internally Displaced and Stateless Persons.

¹²³ See "Al menos 380.863 personas fueron desplazadas el año pasado en el país, según Codhes," *El Tiempo*, April 22, 2009; "Choque por la cifra de desplazados," *El Tiempo*, April 22, 2009; "Número de desplazados en el mundo se mantiene en 26 millones de personas," *El Tiempo*, May 2, 2009.

¹²⁴ See "Cuántos secuestrados y cuántos desplazados?," *Revista Semana*, May 1, 2009; "El conflicto se ha movido a zonas que no estaban afectadas," *Revista Semana*, September 22, 2009.

indicators are seldom used by the media as a description of a humanitarian crisis, in and of itself. In this sense, the role of indicators as a useful instrument to lend notoriety to an issue that is difficult to grasp should, perhaps, be qualified. While indicators have indeed been useful in Colombia to give shape and traction to the IDP crisis in the mainstream agenda, the interest of the media seems to be often mediated by the use of indicators by public officials. This is, of course, not an absolute rule; as discussed in the next section, indicators have become a platform for social mobilization, despite governmental pressures. And yet, most news coverage is not concerned with the indicator in itself, but with the use given to it by public officials, who ultimately remain in control of the agenda.

The connection between indicators and the news media is crucial to understand the former's role as technologies of global governance. Indicators become relevant and influential as news media begin to quote them and use them as evidence of a social reality. To be sure, this move empowers those producing the indicators. Moreover, the role of the press also seems to underscore the dynamics of local use of the indicator. The domestic impact of international indicators is correlated with their prominence in the national press, notwithstanding the fact that all of these indicators, if existent at all, are produced locally and then picked up internationally. This conclusion, though, should be placed in context, as it presumes media that are not controlled by government or other specific interest. What happens if that is not the case? The process whereby indicators become technologies of global governance in contexts with no free press is an important line of further research, which has not been developed yet.

Finally, it is noteworthy that many of these international sources are cautious with their use of these indicators in the Colombian context, always explicitly citing the discrepancies in numbers and the different variations. Domestic media seem uninterested by these subtleties: indicators are appropriated by the domestic media and attributed to international sources, perhaps to give a feeling of impartiality. This closes an important circle; one indicator produced domestically is entered into a process of transformation by being cited by an international report, and then cited back by domestic media. Once the indicator "bounces" back, it acquires a new form, as if it had been produced by a different entity and a different method, and gains renewed authority in the domestic political debate.

IDP indicators as social mobilization

Indicators have become a valuable instrument for civil society organizations active in IDP policy. Most clearly, as we have seen, indicators are useful to place the IDP agenda in the media, communicating a sense of urgency that seems useful in lobbying efforts. Moreover, indicators provide a clear standard for "naming and shaming" domestic governments in their treatment of the IDP situation. Beyond these aspects, though, IDP indicators seem to be more than a tool for social mobilization: they seem to *be* a process of social mobilization, in and of themselves. When doing research for this paper, I was often referred to the Human Rights and Displacement Consultancy (CODHES), an NGO that claims to "represent neither

displaced population, nor the organizations of IDPs."¹²⁵ Rather, CODHES seeks to "prevent the causes of displacement," "evidence the phenomenon of displacement," "analyze the forms of institutional intervention in the displacement,"¹²⁶ etc. All in all, instead of an activist NGO, CODHES seeks to fashion itself as a research institute, a technical source of expertise on IDPs.

CODHES is one of the most important sources of IDP indicators in Colombia. As a private, non-governmental organization, CODHES has been collecting information for the creation of indicators since 1992 through its Human Rights and Forced Displacement Information System, which gathers information through "a systematic and uninterrupted exercise of statistical estimation of contrasted sources."¹²⁷ Since 1995, CODHES manages the Information System on Forced Displacement and Human Rights (SISDHES, for its initials in Spanish), whose methodology involves:¹²⁸ (a) the daily monitoring of more than 450 primary sources (including some international sources) regarding armed conflict and forced migration; (b) reports from other NGOs, some public institutions (such as the Office of the Ombudsman), academic institutions, and the Catholic Church; (c) raw data collected by CODHES staff sent to perform interviews on the ground. Finally, from 1995 to 2004, SISDHES applied the National Poll of Displaced Homes and, since 2007, it supports the Monitoring Commission of Forced Displacement Public Policy in the implementation of a National Poll of Verification, which was put together in the context of the follow-up process to decision T-025 of 2004, in order to apply the indicators the Court had accepted in Writs 109 and 233 of 2007.¹²⁹

CODHES creates the indicators that are then taken up by the UNHCR or the IDMC, and then fed back to the Colombian media, becoming a crucial argument in the domestic debate on IDPs. In this role, CODHES has a fairly activist agenda. The organization is not shy in taking a stand against armed actors and the Colombian government in human rights issues, joining hands with other human rights NGOs. Most of its work is, in fact, along the lines of human rights activism.¹³⁰ My own experience with CODHES gave me that impression. When discussing some of the issues in this paper, I was in contact with CODHES director, Marco Romero. A man of strong character, Mr Romero does not come across as wanting to sound neutral or academic. On the contrary, his interventions often label the administration as a "massive violator of human rights," pointing out the social injustices suffered by IDPs, and their stigmatization in domestic politics. I shared a panel with him in Bogotá at the end of April 2010, where some of the

¹²⁵ See <http://www.codhes.org/index.php?option=com_content&task=view&id=5&Itemid=32>.

¹²⁶ Ibid.

¹²⁷ CODHES, *Cifras e Indicadores del Desplazamiento Forzado y Derechos Humanos en Colombia*, Documentos CODHES No. 3 (2005), 67.

¹²⁸ CODHES, "CODHES Informa" No. 75 (2009), 1.

¹²⁹ Comisión de Seguimiento a la Política Pública sobre Desplazamiento Forzado, *Primer Informe de Verificación Presentado a la Corte Constitucional* (2008), 15.

¹³⁰ A good glimpse is its most representative publications, available at <http://www.codhes.org/index.php?option=com_content&task=blogcategory&id=32&Itemid=46>.

ideas of this chapter were discussed. I can report that, to any observer, there was hardly room for doubt: his is the voice of an activist—and quite a convincing voice at that. In fact, Mr Romero was attacked with a knife one month after we shared that Panel.¹³¹ Fortunately, the wounds (which were apparently the result of a random mugging which had gone wrong) were not mortal. However, as soon as news of the attack spread, the press office of CODHES also released the information that, in the week before, Mr Romero had received death threats from three different paramilitary groups.¹³² News desk EFE suggested that the attack could have been connected to Romero's criticism of remarks by the (then) Minister of Interior, who hinted that IDP funding was being ill-used, as it was being channelled to people who were not displaced, but merely taking advantage of the subsidies.¹³³

In the context of this difficult situation, CODHES spends precious resources in producing all sorts of indicators. Why does it do that? Because such is the language that structures their mobilization. And the strategy is certainly effective, if we understand effectiveness in this context as the marginal increase of the NGO's influence. In its landmark decision, the Constitutional Court called for the creation of a Monitoring Commission of Forced Displacement Public Policy to follow the implementation of policy on IDPs, of which CODHES is a member.¹³⁴ Through its intervention in indicators, this organization has gained access and influence as an expert in the measurement of the IDP situation—an expertise highly valued by the Constitutional Court, as we have seen before. In this way, CODHES has gone beyond producing quantitative data for other activists to use. CODHES does activism by producing the indicators, feeding them with data from different sources (including its own), and then confronting the national government with the large discrepancies between different indicators.¹³⁵ Paradoxically, this move may in turn have demobilizing effects. In the Colombian case, the Constitutional Court opened spaces for public participation in the formulation of IDP policies. While several NGOs used this opportunity, the very complexity of the terms in which the matter had been framed (among other reasons) may have left most of the actual displaced population outside the debate. According to the Monitoring Commission of Forced Displacement Public Policy, less than one-third of IDP organizations are in contact with the national table, where the policy debates take place. In turn, only

¹³¹ See "Director de Principal ONG Defensora de Desplazados fue Atacado por Desconocidos," *El Espectador*, May 23, 2010, available at <<http://www.elespectador.com/noticias/judicial/articulo-204635-director-de-principal-ong-defensora-de-desplazados-fue-atacado-des>>.

¹³² According to the press communiqué, the threats came from "Aguilas Negras," "Los Rastrojos," and "Bloque Capital." See CODHES, "Ataque contra director de CODHES, Marco Romero Silva," May 23, 2010, available at <http://www.codhes.org/index.php?option=com_content&task=view&id=822>.

¹³³ See EFE, "Director de Codhes es herido en presunto atentado" (May 23, 2010), available at <<http://www.terra.com.co/noticias/articulo/html/acu31735-director-de-codhes-es-herido-en-presunto-atentado.htm>>.

¹³⁴ See Comisión de Seguimiento a la Política Pública sobre Desplazamiento Forzado, *Primer Informe de Verificación Presentado a la Corte Constitucional* (2008), 14.

¹³⁵ This is seen through the media coverage of the public debate that arose after CODHES reported almost twice as many internally displaced persons as the Social Solidarity Network. See "Agarrón por las cifras del desplazamiento," *El Tiempo*, September 3, 2004, 3A.

32 percent of all IDPs belong to an IDP organization.¹³⁶ Indicators as social mobilization may, in the end, not necessarily imply grassroots mobilization of the excluded, but rather a high impact strategy of NGOs holding specific technical expertise.

Beyond the influence of a particular NGO, a different standard of effectiveness refers to the impact of the indicators they produce, in terms of changing social policy. In this case, it is hard to assess the impact of CODHES as an independent actor, for its impact is inextricably linked to the emphasis placed by the Constitutional Court on indicators, serving as a sounding board for the particular mode of social mobilization chosen by the NGO. Indicators as technologies of global governance work not as single cause of policy reform, but rather as an important part of a more general narrative of social mobilization that triggers change. This process, though, does not occur in a vacuum, but rather influences and is influenced by the institutional structure. The following section explores the role played by indicators in the interaction between the three branches of power in Colombia: who wins and who loses when indicators are deployed?

IDP indicators and the interaction among separate branches of power

The role of IDP indicators in the political economy of institutional design in Colombia can be usefully understood with reference to two wider debates on governance in Latin America. The first debate is that of judicial activism. As has been discussed, the Court's decision implied a number of structural orders, which were bound to have deep impact in Colombian society. This fact, though, left the Court open to the standard critique of judicial activism. Is it undemocratic to have a Court making this kind of structural decision?¹³⁷ Are Courts the best actor to push for social change?¹³⁸ Does one need to have a deliberative process (ideally through Congress)?¹³⁹ Do Courts have the institutional muscle to enforce this kind of structural order?¹⁴⁰ This discussion exceeds the scope of this chapter. However, the Court clearly faces severe challenges in justifying its ambitious intervention. There is an underlying problem of legitimacy that cannot be ignored.¹⁴¹ It is not as if the Court can simply give an order, and all the bureaucracies of the Colombian

¹³⁶ Comisión de Seguimiento a la Política Pública sobre Desplazamiento Forzado, "Primer Informe de Verificación Presentado a la Corte Constitucional" 2008, 49.

¹³⁷ See Jeremy Waldron, "The Core Case Against Judicial Review," *Yale Law Journal* 115 (2006): 1346.

¹³⁸ See Ran Hirschl, *Toward Juristocracy: The Origins and Consequences of the New Constitutionalism* (Cambridge, MA and London: Harvard Press University, 2004), 169.

¹³⁹ See, e.g., the notion of populist constitutional law in Mark Tushnet, *Taking the Constitution Away from the Courts* (United Kingdom: Princeton University Press, 1999), 177.

¹⁴⁰ For a summary of this argument (and the counter argument) in the South African context, see Mark Tushnet, *Weak Courts, Strong Rights: Judicial Review and Social Welfare Rights in Comparative Constitutional Law* (New Jersey: Princeton University Press, 2008), 196.

¹⁴¹ The term "legitimacy" is used here in its normative sense, as an answer to the question: "Why should the structural intervention of the Court be obeyed?" In Weberian terms, the Court's intervention needs to command uncoerced obedience. To that effect, as we will see, indicators are useful. On the normative/empirical aspects of legitimacy in global governance, see Dencho Georgiev, "Politics or

establishment happily bow to its commands. On the contrary, one finds bureaucratic inertia, turf wars, and lack of will. While throughout the 2004 decision (and the follow-up process), one notes the Court's constant concern for respecting the sphere of action of the institutions involved with IDP policy, the fact is that the sheer scope of the orders was bound to create a certain backlash from the administration.

This is, to be sure, not the first time that one of the Court's decisions sparked severe debate on the legitimacy of its far-reaching decisions.¹⁴² While the Court's objective seems to have been to jump-start lethargic bureaucracies and bring them to speak the language of rights, it still shows a certain restraint that can be understood as deference to other branches of power. There is a constant need to legitimize (and re-legitimize) the structural decision taken by the judge, as opposed to the same decision taken by elected representatives in Congress. Enter thus indicators. Indicators provide the Court with at least four valuable assets in its quest for legitimacy:

- (a) First, they provide objective criteria to evaluate the reasons for the Court's structural intervention. In much the same way as indicators provided the UNHCR with rationality and predictability in its involvement with IDPs' situations, the same technology provides rationality to an intervention by the Court that could be perceived as random and unpredictable.
- (b) Second, indicators provide transparency to the Court's structural intervention. While the decision-making process within the Court could be perceived as obscure, indicators set (or at least, give the impression of setting) certain guidelines for decisions that are open and known to all.
- (c) Third, indicators provide a standard to assess whether the Court's structural intervention is needed or not. Ultimately, once the "jump-start effect" has faded away, the longer the Court stays involved with a situation such as the IDP crisis in Colombia, the less legitimacy its intervention has. Indicators are useful in this context, as they provide a reasonable exit strategy to the Court, once its intervention has worn out its welcome.

Rule of Law: Deconstruction and Legitimacy in International Law," *European Journal of International Law* 4 (1993): 12ff.

¹⁴² Prior "unconstitutional state of affairs" included the situation of public funding of education (Decision SU-559 of 1997), the inhuman conditions in Colombian prisons (Decision T-153 of 1998), and the general vulnerability of human rights activists (Decision T-590 of 1998). The most controversial decisions, though, have been those that are perceived to have deep impact on the country's economy. For example, the Court's legitimacy was questioned on the occasion of Decision C-700 of 1999, where it decided that the mortgage and housing financing system in force was contrary to the Constitution. Controversy was also sparked when the Court decided that the yearly salary raise of civil servants should be equivalent, at least, to the yearly consumer price index (Decisions C-1433 of 2001, C-1064 of 2001, and C-1017 of 2003), and when the Court decided that the yearly raise of the minimal wage should be equivalent, at least, to the inflation rate (Decision C-815 de 1999). For a critique of the latter decision from the economics perspective, see Marc Hofstetter, "Política Monetaria y la Corte Constitucional: El Caso del Salario Mínimo," Documento CEDE 2005-36.

- (d) Fourth, the Court's turn to indicators fostered a process of public debate underscoring its legitimacy. As we have seen, the Court called different civil society organizations and other stakeholders to propose and evaluate IDP indicators (with varying degrees of success). The political dynamics triggered by this call, the discursive practices it inspired, seem useful to further advance the legitimacy of the Court's structural intervention. This suggests a different angle for understanding the connection between indicators and legitimacy. One constant preoccupation with the use of indicators as a technology of governance is that, by turning to quantitative data, it may hinder democratic decision-making, instead empowering experts and technocrats. However, the Colombian Court tried to foster social mobilization by inviting organizations to create and criticize indicators, thereby also trying to expand popular support to a judicial intervention that was certain to draw fire. In this case, then, instead of featuring a sharp contrast to democratic legitimacy, the Court's turn to a dialectical construction of indicators can be read as a move to actually increase (again, with more or less success) the decision's legitimacy.

The combination of these four variables contributes to a more convincing justification for the Court's structural intervention in the IDP case. It should be noted that the IDP decision was one of the very first decisions where the Court placed an emphasis on indicators. Since then, quantitative methodologies have become a central tenet when a far-reaching structural decision is on the table. A case in point is a recent decision concerning the structure and financial stability of health care in Colombia.¹⁴³ Perhaps inspired by its experience following the IDP decision, the Court gave structural orders to all institutions involved with health care, and put in place a follow-up process based on writs. Once again, the Court held that an indicator of the "effective enjoyment" of the right to health was required.¹⁴⁴ And, once again, the Court adopted follow-up writs where indicators play an important role.¹⁴⁵ The Court formed an expert committee comprised of deans of medical schools, lawyers, and epidemiological experts. The group was tasked with identifying parameters for creating indicators to evaluate government's compliance with health rights. In its task, the experts group drew on norms and guidance of the World Health Organization, the Pan-American Health Organization, the Inter-American Court of Human Rights, as well as literature by Western academic scholars. In October 2011, the Constitutional Court adopted Writ 226, in which it established the framework for its interaction with Congress, the government, and independent regulators of health care. Indicators are at the core of the Court's approach. Expressly inspired by recent scholarship on indicators as technologies of governance, the Court developed quantitative parameters to be followed by regulators and Congress, under which

¹⁴³ See Decision T-760 of 2008.

¹⁴⁴ See Decision T-760 of 2008, Section 9.

¹⁴⁵ See, e.g., Writ of July 13, 2009 (no number), concerning the implementation of Order 20 in Decision T-760 of 2008.

the Court will monitor their performance.¹⁴⁶ In this sense, the turn to indicators fits well in the wider debate of judicial activism. When a structural intervention is advanced, emphasis on indicators provides a useful platform for the judiciary to enhance the legitimacy of such intervention.

Moreover, the role of IDP indicators in the political economy of institutional design in Colombia can be also usefully understood with reference to a second axis—this time, the general trend towards rationalizing administrative action. The evolution of Colombian administrative law can be understood as a never-ending quest to achieve efficient bureaucracies, where technocrats would populate the administration, and exercise power rationally and predictably. This ideal of expelling politicians and their electoral rationality from the administration has been usefully labeled the “Weberian Dream” of Colombian administrative law.¹⁴⁷ The turn to indicators in the Court’s decision seems to be another expression of this dream. The Court pushes for indicators, because it wishes to empower rational bureaucracies who may be monitored, and who accept the legitimacy of quantitative methods. To be sure, this focus on indicators underscores the relation between the Court and the Executive branch. Congress is left in an accessory role, a guest whose internal dynamics of deliberation is an ill fit for the quantitative language of indicators.

Two different attempts at passing IDP-specific legislation were attempted at the end of the 1990s, both of which successfully culminated in the passing of the aforementioned Law 386 of 1997.¹⁴⁸ Colombian legislative procedure demands that an ordinary bill (one which does not modify the Constitutional Bill of Rights or Congress’s powers) receive two separate debates from each Chamber in their respective committees, and in their plenary sessions. Accordingly, each ordinary bill is discussed four times before its passing, and its author is allowed to deliver a general statement to support the project in its initial debate. Consequently, Law 386 of 1997 had five specific parliamentary discussions: its first general statement, and four reports in the subsequent discussions in Chambers. None of these reports, discussions, or statements focused on IDP indicators. The bill’s authors quoted the amount of displaced persons as reported by the Episcopal Conference, which counted 600,000 displaced persons at that time.¹⁴⁹ However, the general statement took a cautious approach towards the use of indicators, by declaring expressly that “although the registered quantitative variables show great discrepancies and are in themselves not too eloquent of the dimension of the problem, the latest research by the Human Mobility Section of the Colombian Episcopal Conference show the existence of six hundred thousand displaced persons [...]”¹⁵⁰ This was further

¹⁴⁶ See Writ 226 of October 21, 2011.

¹⁴⁷ See Diego López-Medina, “El Sueño Weberiano: Claves para una Comprensión Constitucional de la Estructura Administrativa del Estado Colombiano,” *Revista de Derecho Público-Universidad de Los Andes* (2007): 1.

¹⁴⁸ These attempts were Bill 033 of 1996 and Bill 016 of 1995 of the Chamber of Representatives, and Bill 015 of 1996 of the Senate (which corresponds to the Chamber of Representatives’ latter Bill). Both of these projects were joined in the Senate’s plenary session, and thus, had the same discussions and voting procedures.

¹⁴⁹ See Gaceta del Congreso, Tuesday, August 8, 1995, 2.

¹⁵⁰ Ibid.

stressed when quoting the Colombian Ombudsman, who espoused a similar approach: "statistics of catholic hierarchy show that out of the 24 million of displaced persons worldwide due to non-international armed conflicts, 548,261 are Colombians [...]." ¹⁵¹

The first debate in the Chamber of Representatives featured no reference at all to IDP indicators. This suggests, alongside the debate lacking in plenary sessions, that Congress was not particularly concerned with indicators, perhaps due to executive action on the same issue, which was much more far-reaching than parliamentary action. ¹⁵² The second debate in the Chamber of Representatives, which took place one legislature after its first debate, included only a fleeting reference to any indicator, by an unsubstantiated claim on the total number of displaced persons: "Indicators are growing in an alarming manner, today there are 700,000 Colombians affected and the consequences and effects that this situation has for our nation are intangible." ¹⁵³ The first and second reports in the Senate featured the same report and both failed to be discussed in the plenary. Ultimately, the debate in the lower Chamber was perhaps the most extensive debate on the question of IDPs to have ever taken place in Congress—and there was no reference whatsoever to indicators. And yet, Congress did understand that use of indicators was a discursive strategy available to MPs. In the following year, after the Chamber of Representatives' second debate, the Senate was using a higher number of internally displaced persons and unfounded indicators regarding the compositions of displaced groups: "It is estimated that the number of displaced persons today is of 750,000 people and the most affected groups are those of women and young people. It is considered that women represent 58 percent of the displaced population. [...] In the same vein, 72 percent are younger than 25 years of age." ¹⁵⁴

Indicators *were* part of the landscape, but Parliamentary arguments were not ultimately based on them. Even if quantitative data was sometimes introduced in debates, Congress was not that interested in indicators. This may be counter-intuitive, as one could expect Congress to be interested in limiting the discretion of the executive branch in dealing with the IDP situation, and implementing a system of bench-marking based on indicators may be perceived as a way of doing so. However, indicators may be perceived as an expertise held mainly by executive agencies. Therefore, it is possible to interpret Congress' systematic rejection of indicators as a tool of governance as a way of keeping its control over the IDP issue.

In contrast, the Court seemed clearly interested in enhancing the use of indicators. Indicators can be read, therefore, as a part of a more general agenda of good governance, in the sense that they would be perceived as enhancing accountability, transparency, and rationality. This move also empowers technical experts, who have knowledge on indicators, and could be read as disempowering Congress, who is

¹⁵¹ See *Gaceta del Congreso*, Tuesday, August 8, 1995, 3.

¹⁵² As shown before, during the 1990s and up to Decision T-025 of the Constitutional Court, normative efforts to tackle IDPs were primarily of an executive nature. These included the above-mentioned Conpes documents and presidential decrees.

¹⁵³ *Gaceta del Congreso*, Tuesday, March 12, 1996, 3.

¹⁵⁴ *Gaceta del Congreso*, Thursday, June 5, 1997, 3.

also disempowered if the “Weberian dream” becomes a reality. In the Colombian case, the Constitutional Court seemed to put its weight on the side of the technocrats, as perhaps this move would also empower the Court to control the discussion on IDP policy, by framing it as a matter of indicators and monitoring, and not as a problem of redistribution of wealth, which would require a wider debate in Congress.

Paradoxically, for all its tense relations with dormant bureaucracies, the judiciary is brought closer to executive agencies by the turn to indicators. Indicators work as a common language between the technocrats in government and the lawyers in the Courts, which allows a certain communication that would have impossible otherwise. Through this common language, and particularly through the concept of indicators of “effective enjoyment of rights,” the Court gets involved in the logic of governance—distributive choices, second bests, Catch-22 situations. Bureaucrats, in turn, get involved in a logic of rights that accepts the counter-majoritarian rationale of adjudication. This process, in turn, has one final interesting feature: the turn to indicators seems to create middle ground between lawyers and economists, where the disciplinary loyalties of each can be fruitfully negotiated.¹⁵⁵ As such, then, indicators can be usefully understood as a technology of governance that crosses both national and disciplinary borders. And yet, Congress seems to be, at any rate, left out of the party.

Conclusion

The IDP experience in Colombia points to some promising aspects of researching indicators as technologies of global governance. It also reveals some limitations in the understanding of the problem up to now. It shows that indicators do indeed spark an interaction between global decision-makers and domestic agencies. This process is one in which the indicator gains a life of its own in the domestic political debate. In that sense, indicators have influence in domestic decision-making and are indeed a mechanism for exercising global authority. And yet, there remains a space for use and misuse of the indicator, where global agendas get transformed and become the stuff of domestic politics. In those contexts, indicators serve unexpected goals. In the IDP case, for example, they became the sole end of social mobilization, and became part of century-long debates on judicial activism and administrative transformation.

How can the specific influence of indicators in domestic contexts be assessed? While it seems hard to establish a test of strict causation, the IDP experience in Colombia shows that introducing indicators into domestic debate changes the way in which issues are understood and spoken of, empowering some above others. The

¹⁵⁵ For the wider implications of the struggle between lawyers and economists in Latin America, see Yves Dezalay and Bryant Garth, *The Internationalization of Palace Wars: Lawyers, Economists, and the Contest to Transform Latin-American States* (Chicago and London: The University of Chicago Press, 2002).

lesson seems to be that we need a more robust theory of global influence and authority in order to determine the specific ways in which indicators change the decision-making process. It also shows the limited use of the conceptual structure featured in previous work on indicators, where the variables in the triad of global governance are governors, governed, and the public. Once the indicator enters the domestic debate, the "public" will use it, but so will the "governed." Once again, the directionality of authority is less simple than it may appear. As for the "governors," the IDP experience shows that they are both users of, and measured by, indicators. The UNHCR itself needs to justify its intervention in IDP contexts, and indicators come in handy to that effect.

Indicators are a representation of reality. However, in a way, they create a reality as well. The IDP debate in Colombia is a demonstration of this. While the marginal effect of violence made it difficult to grasp the magnitude of the human tragedy, the introduction of indicators to the debate made it possible to name that reality, and connect a number of circumstances (poverty, violence against women, rising prostitution, and suicide rates) with a single cause (displacement due to internal armed conflict). The role of news media in that sense is of crucial importance; it is, quite literally, the mediator between the technical expertise embedded in the statistic and the political happening of the indicator as a technology of global governance. In much the same way as mere statistical data is not an indicator, a mere indicator is not a technology of governance without the media.

Ultimately, the case study featured here confirms the idea that indicators rapidly gain a life of their own. This life is truly global, as it involves both an international and a domestic (even local) aspect. As hinted at before, it is by looking at the intimacies of its domestic life that we get a better sense of the inner working of this technology of governance, the reasons behind its success, and the specific challenges it poses.

Problems of Power in the Design of Indicators of Safety and Justice in the Global South

*Christopher Stone**

In matters of crime, safety, and justice, indicators are instruments of power. They are not merely this, for indicators can also be sources of insight and pride, promoting good governance of law enforcement and criminal justice through inspiration rather than coercion. Still, the reason for the present interest in these governance indicators is their utility for those who would exercise power.

In other walks of life, indicators play a more straightforward role. Oral temperature and blood pressure are relatively benign indicators of personal health, just as barometric pressure is an uncomplicated indicator of weather. However, the ranking of cities by their homicide rates, the discussion of low rates of criminal convictions, or the publication of surveys reporting rising fear of crime are not mere diagnostic exercises; they are efforts to shape in particular ways the priorities of police, prosecutors, judges, and their respective ministries and departments. They are efforts to turn the legal and penal apparatus of the state to particular purposes.

Indicators, as defined in the introductory chapter in this volume, are named collections of data, simplified and processed into measures of the performance of some organizational unit. As that same chapter makes clear, indicators are beset with problems: their names usually promise more than measurement can deliver, their scientific covering is too thin to provide much comfort, and their movements are difficult to interpret. In the case of indicators of safety and justice, the organizational unit whose performance is being measured is itself problematic because it rarely corresponds to how the criminal justice system (or the “justice sector”) is in fact organized. Still, despite all these challenges, indicators of safety and justice are useful tools in the hands of those who would strengthen the rule of law, reduce crime, and administer justice. The question is not whether indicators, despite their flaws, can help to achieve these worthy ends. They surely can; the question is rather whose vision of safety and justice they advance.

As instruments of power, indicators are available to many different actors with various relationships to formal authority. Elected officials can employ indicators to

* An earlier version of this chapter was presented as a paper at the conference on Indicators as a Technology of Global Governance, held at NYU Law School, September 13–14, 2010.

manage operations that they directly oversee—a straightforward use of indicators as a management tool to exercise legitimate authority. Where, however, multiple officials at different hierarchical levels and in different departments are vying for control over some government policy, and where NGOs, foreign governments, and international bodies are also eager to influence this policy, advancing a particular set of indicators can be a subtle yet effective way of gaining some control of government policy and priorities. This is particularly so when the more blatant exercise of power might appear to be illegitimate. What makes indicators attractive in these circumstances is that they carry their own technical claims to legitimacy.

The use of indicators as instruments of power is particularly striking in developing countries where international donors and institutions of global governance exert considerable influence. A national police service may find its performance measured by dozens of competing indicators specified by its commander, by the country's prime minister, by the finance ministry, by multiple bilateral donors, and by various agencies in the UN family. In many instances, none of the indicators are used effectively, nor are the data collected accurately. The problem may appear to be a lack of skills, inadequate equipment, or nearly constant requests for new indicators. No sooner is a data system purchased to calculate one set of indicators than the clerks are directed to find the data needed for a different set. The result, as one aid official put it to me, is that many developing countries are littered with the carcasses of failed indicator projects—the consultants paid and gone, and those charged with administering justice increasingly cynical about time wasted on measurement when there is real work to be done.

In these settings, however, a field of failed indicator projects may be a sign, not of skill deficiencies, but of resistance in a particular agency to the exercise of power from some other government department or from abroad. The challenge for those ambitious to advance the use of indicators across government or internationally—whether in a prime minister's office or an international development agency—is to align indicators with the ambitions of those in positions of immediate and legitimate authority.

It will help here to consider indicators deployed at three distinct levels of governance: within an individual ministry or government department, across government as a whole, and at the level of global governance. The same indicator might be used at all three levels, but the indicator exerts power on behalf of different people and different purposes at each level. For example, a police chief might use changes in survey results measuring public fear of crime in various neighborhoods to reward or demote commanders of the corresponding divisions, while a governor or treasury official might use the same measure of public fear on a government-wide basis to spur better coordination among police and prosecutors. At the level of global governance, the UN, the World Bank, or a bilateral aid agency might use the same indicator of fear to compare the effectiveness of national governance across several countries or to measure the relative success of their own projects in a region.

Rarely, however, are the same indicators used at all three levels. More commonly, a police chief will rely on one set of indicators, including the numbers of arrests, the

amount of contraband seized, and the level of public support for the police revealed in occasional surveys. The governor may focus on the numbers of reported homicides and robberies, the level of fear among the public, and the cost of police services. At the same time, various international bodies may be rating the same jurisdiction on its prosecution of human trafficking cases, the perception of corruption in the police, and the speed with which it completes trials in criminal cases. The resulting cacophony of indicators is not simply the result of the complexity of efforts to deliver both safety and justice, but a result, too, of competing priorities and competing strategies at different levels of authority over the issues of crime and the administration of justice.

The indicators at these three levels interact. At each level, it is tempting for those exercising power to imagine that they can design and deploy indicators unilaterally, as if other indicators and their sponsors were not simultaneously seeking to shape the performance of the same government system. But for those on the receiving end of these indicators—such as the police, prosecutors, magistrates, corrections officers, clerks, and others engaged in making the rule of law real—the competition among indicators is unavoidable. The conflicting incentives, partial understandings, and ideological assumptions in the various sets of indicators compete not only with one another, but also with the values and ambitions of those doing the work.

This essay explores the possibility that governance indicators can be harmonized across these three levels, and that doing so will produce effective governance. Specifically, I argue that those operating on the global level—particularly in the domain of safety, justice, and the rule of law—might design their indicators from the bottom up, supporting local ambitions and building on the legitimate sources of authority close to the operations they seek to influence, rather than starting with ambitions and power at a global level. The chapter draws on work my colleagues and I are conducting with government officials and civil society leaders in Jamaica, Nigeria, Papua New Guinea, and Sierra Leone, sponsored by the British Department for International Development (DFID) and the Australian Agency for International Development (AusAID). I draw as well on a separate effort led by the Altus Global Alliance and supported by DFID that has been taking a similar approach with indicators of police reform in many countries across Asia, Latin America, and Africa.

Starting with local ambitions and local, legitimate authority may help repair a persistent problem in the construction of indicators for global governance, namely the weak sources of legitimacy for the standards implicit in indicators at the global level. If by starting at a local level, our partnerships can produce global indicators grounded in local standards and ambitions, we might forge a stronger global solidarity in the effort to improve the quality of justice and strengthen the rule of law.

I argue here that such a bottom-up approach is not only possible and practical, but has the potential to engage citizens and domestic leaders enthusiastically in a creative and democratic construction of justice. In the next section, I describe what such indicators look like when designed collaboratively with domestic government officials responsible for some part of the administration of justice. I introduce the

term “active indicators” to distinguish these indicators from those designed without the participation of local, operational authorities. In the third section, I describe how active indicators can also be created for use by citizens in civil society operating with a less formal, but nonetheless legitimate, authority. Here I also begin to sketch how locally owned indicators can be aggregated up to a global level. In the fourth section, I return to the questions of how active indicators built at the domestic level might be spread from country to country, eventually creating a coherent, global set of common indicators.

Active indicators of safety and justice

On a hot day in February 2010, the Commissioner of the Jamaican Constabulary Force walked into a meeting of his command staff, already in progress. He stood in the back of the darkened room, catching a few minutes of a presentation to his staff on “change management.” Before the discussion period even began, he turned and left, but not before telling the meeting organizer from the Ministry of National Security that he would be rearranging his schedule to get back at the end of the day—long after the meetings had ended—to talk more about the presentation. He had seen something he could use.

Relatively new in the top job of his long-troubled national police service and not yet the master of his schedule, the Commissioner returned even later than he had planned. He took a chair in the now nearly empty conference room and explained what had lured him back. He wanted to see again a scatter-plot that my colleagues had constructed with his staff and with a researcher in the Ministry of National Security, showing the number of raids, vehicle searches, and pedestrian stops that each of his 19 police districts had conducted in 2009, along with the number of “hits” each had recorded during the year: offenders arrested and guns and vehicles seized (see Figure 11.1). The scatter-plot revealed that for most of the divisions, there was a straightforward relationship: the more raids and searches, the more hits. But a handful of divisions were worryingly apart from the others on the chart, having conducted many more raids and searches but achieving no more hits than divisions that had conducted a tenth as many. More encouragingly, there were one or two divisions that had recorded high numbers of hits while apparently conducting relatively few searches and raids.

The Commissioner had seen all that in an instant during the morning presentation and had returned because he wanted to use the chart to better understand the apparently good work of the highest achieving division. If the results were real, he wanted the commander of that division to explain to his counterparts at an upcoming command meeting what he was doing to get those good results while intruding relatively infrequently on the lives and liberties of local residents.

The Commissioner’s interest spurred his staff and the team that had created the scatter-plot to dig down into these data, attempting to chart the trends month by month and testing which reports were reliable and which spurious. The team had to harmonize inconsistent definitions of searches and seizures, and facilitate agreement

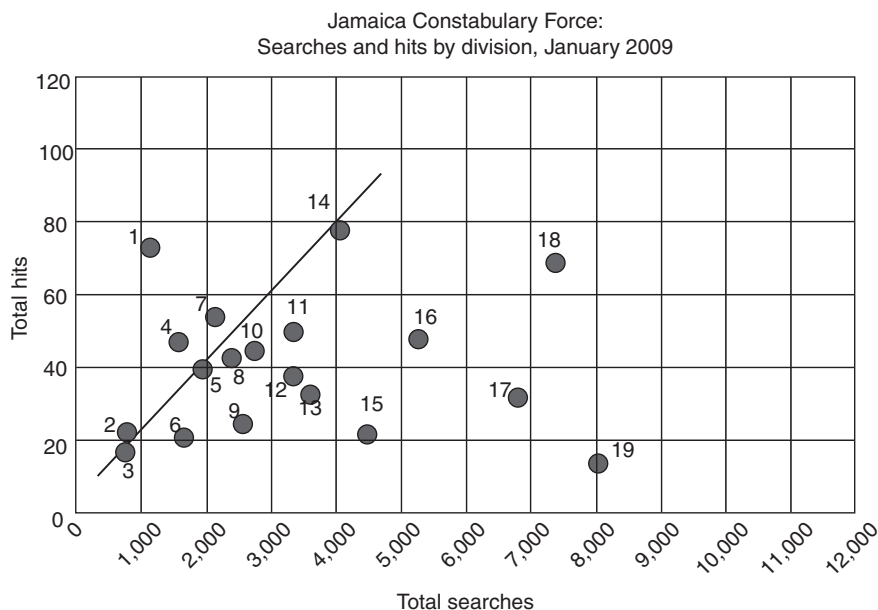


Fig. 11.1

Source: Program in Criminal Justice Policy & Management, Harvard Kennedy School
Based on data supplied by the Jamaican Constabulary Force, 2010.

on how to count searches and hits. In May 2010, the meeting that the Commissioner had envisioned took place, with all of the divisional commanders discussing a scatter-plot that showed how each division had moved both on “searches” and “hits” from January through March 2010.

The Commissioner saw the power in these indicators, and restructured relationships so that the head of the unit devising these measures reported directly to him. Then, over the next 18 months, the unit trained the station officers-in-charge, divisional data clerks, and headquarters analysts in the continued production of this indicator, and refined the measurement tools so the Commissioner and the division commanders could use monthly scatter-plots on a continuing basis to improve the effectiveness of the raids and searches they conduct.

The relationship of hits to searches, illustrated on the Jamaican scatter-plot, is an example of an *active indicator*. It is designed specifically for use by an official with formal authority over the people expected to produce the outcome being measured. In concrete terms, active indicators: (i) capture performance in tight timeframes: usually days, weeks, or months; (ii) present data at the level of operational responsibility; (iii) ground discussion at management meetings where officials are accountable for results; and (iv) describe outcomes in common language, often in graphic form, understandable by people both inside and outside the institution.

A few weeks later on the other side of the globe, some colleagues and I sat with a wise but dispirited assistant commissioner of the Royal Papua New Guinea

Constabulary. Taped to the wall of the conference room in police headquarters in Port Moresby was a page from a flip-chart titled "Strategic Planning," apparently from a meeting some days earlier. A single person's handwriting filled the large page with a couple of dozen generic categories used by consultants everywhere, from "vision and mission" at the top to "monitoring and evaluation" at the very bottom, almost falling off the page.

The assistant commissioner explained that he was in charge of corporate services and administration, including all of the Constabulary's monitoring and evaluation, but his department had been starved of resources and kept apart from the operational core of the organization. As a result, he has little up-to-date performance data. Observing an operations management meeting for the National Capital District a few days later confirmed the assistant commissioner's description. Vehicle thefts are on the rise, said one participant; but this was a data-free meeting, with no evidence of trends farther back than the latest anecdote.

Until recently, demand for quantitative indicators of police performance in Papua New Guinea emanated from outside the RPNGC. In the Law and Justice Sector Secretariat, the "monitoring and evaluation unit" is responsible for reporting on 64 indicators annually, and does so typically about six months after the close of each year. The report for 2009—the sixth Annual Report—includes relatively little data collected from the Constabulary, but instead relies heavily on a Community Crime Survey that shows "many people . . . reporting poorer perceptions of police" and higher levels of crime victimization.

The indicators in the Annual Report are not referenced in the operational meetings of the Constabulary and the reason for the distance between them is readily apparent. The Annual Report is the product of the Law and Justice Sector Secretariat, which—according to the opening pages of the report itself—was established to support the Law and Justice Sector Working Group, which was established to support the National Coordinating Mechanism, which is overseen by the Ministerial Committee on Justice, Law and Order. This nested set of coordinating bodies is charged with implementing the "three pillars" of the National Law and Justice Policy and Plan of Action, and has advanced that project by developing a Sector Strategic Framework with five goals. This, in turn, has given rise to a Performance Monitoring Framework with 18 indicators and 64 "sub-indicators" to track progress toward the five goals. These are not active indicators.

The Jamaicans have their equivalent of the 64 indicators in Papua New Guinea, as does just about every nation, province, and municipality at every income level. They can be useful, but they are not powerful: they are not suited for active management but—at best—for more general diagnosis. Their architects hope that they can align the long-term goals of separate government departments with each other and, equally important in developing countries, with the goals of international donors. Such indicators may provoke conversation and they may inspire reports in the press, but on their own they rarely spur action in operational agencies. They represent a tentative consensus about a collection of more or less compatible ambitions, but they do not have the force of fully determined authority.

The Commissioner of the Royal Papua New Guinea Constabulary is a member of the National Coordinating Mechanism, and so has a stake in the 64 indicators. He is reportedly pleased when they imply that the police are doing something right, but these are not the indicators he needs in order to actively drive change in the police. When the assistant commissioner for RPNGC's corporate services saw the scatter-plot designed for the Jamaican Commissioner, he recognized the difference immediately and began to think about how he could create indicators for his Commissioner that would be similarly useful for month-on-month management of the police.

In addition to serving the needs of a single departmental leader, active indicators can be built to support management across separate government institutions. Consider, for example, the issue of pretrial detention in Lagos, Nigeria. In any country, pretrial detention can easily escape management attention because the problems of prolonged detention and overcrowded conditions are felt in remand prisons, yet prison superintendents have no authority to solve the problems. They cannot shorten the stay of detainees or prevent new ones from arriving. Those who possess that authority—prosecutors and judges—do not see or experience the problem of overcrowding on a day-to-day basis, allowing them easily to overlook it. In Lagos, that universal problem is compounded by Nigeria's federal structure. Prisons are operated by the federal government and the police who prosecute minor cases are federal employees, yet prosecutors in serious cases and the criminal courts themselves are under the control of each state. When the Attorney General of Lagos State became concerned about the problem of pretrial detention, he understood the jurisdictional complexity and so joined forces with the prison administration to create an active indicator that would allow him and prison officials to begin to manage the problem in meetings with prosecutors and the judges.

Why did the Attorney General need to create a new indicator? A simple search on the internet would have told him that 65 percent of the prisoners in Nigeria were being held pretrial. Indeed, this indicator is available for almost every country in the world. Why did the Attorney General need anything else?

The percentage of prisoners in pretrial detention is not an active indicator. No one is accountable for raising or lowering it, nor would a reduction in the number necessarily mean that the number of people in detention was being reduced or that overcrowding was being eased. If, for example, prison sentences were substantially lengthened, the percentage of people in pretrial detention would fall, without reducing their numbers. Indeed, overcrowding would become worse.

Nor could the Attorney General simply use as his indicator the number of people in pretrial detention. This indicator would have two faults. First, the raw number in detention is the product of so many actions by people in so many different institutions—police, prosecutors, judges, and legislators—that no one feels responsibility for it. Second, pretrial detention itself is not unjust. It only creates injustice when (a) it lasts excessively long; (b) it is imposed on people accused of trivial crimes; (c) it is imposed on people without even minimal evidence of their guilt; or (d) it entails inhumane conditions of detention. An active indicator of pretrial detention must focus the attention of officials on one or more of these four specific

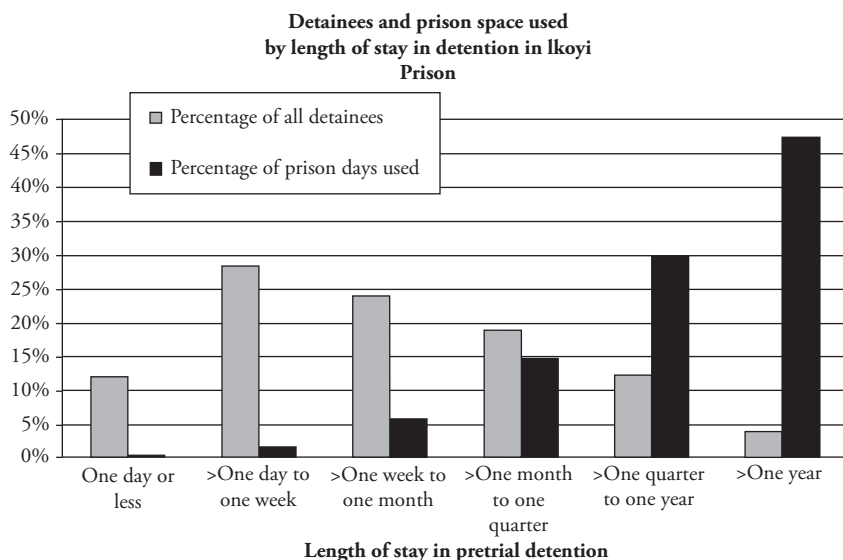
problems in their direct control. Because so many of the horror stories of pretrial detention in Nigeria concerned people detained for years without trial, the Attorney General decided to focus on problem (a): the length of pretrial detention. Now he needed an active indicator.

One only knows the full length of a suspect's pretrial detention when it ends. So to build an indicator of the length of detention, the Attorney General's staff, with the assistance of the Lagos-based CLEEN Foundation and some of my colleagues and students, assembled a sample of exit data from early 2010 already collected manually by the prisons in Lagos State and available in giant ledgers. The data turned out to include not only how long each person had been detained, but also the court in which the case was heard, and how each person's pretrial detention ended: with conviction and sentence, dismissal of the case, payment of bail, or some other event.

The first thing the exit data taught everyone involved was how many suspects were already staying a very short period of time in detention. So much attention had been focused on the long-term detainees that the hundreds of prisoners cycling in and out of detention were almost invisible from a prison management perspective. Both prison staff and prosecutors were therefore surprised when the data revealed that the median length of stay for all detainees in the 2010 sample was 14 days. Looked at in more detail, the data on length-of-stay in detention revealed that the officials in Ikoyi Prison were facing the same double challenge the superintendents of most remand prisons face around the world: most of their detainees remained a very short time in detention, but the relatively few detainees who remain a long time use up the lion's share of the prison resources. Specifically, the 2010 sample from Ikoyi Prison showed that the suspects remaining in detention a month or less accounted for almost two-thirds of all detainees (64 percent), but accounted for only about 8 percent of the prison space occupied over time. In contrast, only about 4 percent of suspects remained in detention for a year or more, but accounted for almost half (47.5 percent) of the prison space occupied over time.

Working from the existing data, it is possible to construct an active indicator that might allow the Attorney General and the prison officials in Lagos to manage the size of their remand prison population. Reducing each stay by one week would reduce the prison population by 8 percent, a worthwhile strategy, since the target seems achievable and most of the suspects are staying a short time in any case. At the same time, completing all of the long-term cases (those that have already kept a suspect in detention for a year) before the end of another six months in detention would reduce the remand population by a further 17 percent, for a total reduction of 25 percent. The first strategy would have to be applied to all cases, while the second could be concentrated on the courts handling the long-term cases, which account for only 4 percent of the detainees. If these strategies succeed, the black bars on the chart in Figure 11.2 should begin to even out, while the grey bars should be concentrated even more on the shorter stays.

The two indicators from Jamaica and Nigeria are “active” not only because they are sensitive to interventions, measured in tight time frames, expressed in simple language, and designed for particular management activities, but also because they

**Fig. 11.2**

Source: Program in Criminal Justice Policy & Management, Harvard Kennedy School
Based on data collected in Ikoyi Prison, 2010.

are linked to particular officials who exercise legitimate authority over the subjects being measured. The divisional commanders in the Jamaican Constabulary Force are clear about the Commissioner's authority, so the indicator relating searches to hits can effectively serve the Commissioner's purpose. In Nigeria, the Lagos State Attorney General has made common cause with the superintendent of the prisons in his state, and between them they can effectively push the Director of Public Prosecutions and the courts to help them with a problem visible principally in the prisons. If there is a weakness in the pretrial detention indicator, it is the weakness of the link to the police who, as the prosecutors in minor cases, also contribute to delays.

Ideally, active indicators at a department or sector level are aligned with the national, long-term, and mid-term plans; but such alignment should not be taken for granted. It is common to hear complaints from officials in operational agencies that they are measured against too many different sets of indicators: their own boss has one, the finance ministry has another, the President or Prime Minister may have a third, and international donors have a handful of their own. The multiplicity of indicators is not, in itself, a problem if they are well and clearly aligned, but too often they are developed independently of one another, overlapping only by coincidence, competing for the attention of front-line officials increasingly cynical about measurement.

It is difficult, perhaps impossible, to get meaningful commitment to an aligned set of indicators by starting at the global level. The officials and consultants

developing the indicators at that level rarely have the time or opportunity to dig down into the operational units of the justice sector and understand the priorities driving performance there. Relying simply on the formal agreement of domestic officials to indicators developed at the global level will rarely produce either alignment or commitment. Such agreements are politely made and easily ignored. These two examples from Jamaica and Nigeria suggest a different course: fashioning indicators first to advance the strategic ambitions management purposes at the domestic level, and building up from there.

What would such a program look like? In Papua New Guinea, for example, a working group that spans several operational departments is building active indicators for the coordination of customary and magistrates courts and for the management of private security services. The effort is coordinated by the Law and Justice Sector Secretariat and the Department of National Planning in order to maintain explicit links among the active indicators, a new set of mid-term development indicators, and the longer standing 64 sector-wide development indicators. These design and alignment processes recognize that the authority that these indicators most directly convey is the power of those leading the operational agencies.

Indicators as a source of power for civil society

Active indicators, by definition, are tightly linked to authority, but this need not be the authority of government officials. Active indicators can serve citizens and civil society as well as the state, for there is legitimate authority in civil society, too.

Consider the common need to improve the quality of operations and service at police stations. This might include the good functioning of equipment, the knowledge of police officers in each station, and the conditions of detention for suspects. It would probably also include the transparency and accessibility of those stations to the people in the districts they serve, specifically the clarity of signage, the display of useful public information, the provision of waiting rooms, officers who do not expect a bribe to accept a report, and the courtesy and respect with which those officers meet visitors.

There is nothing to stop senior police officers from creating active indicators to monitor all of these aspects of the quality of their police stations, but such officials usually put a higher priority on issues of enforcement activity and discipline. Citizens themselves, in contrast, tend to put a high priority on these “customer service” aspects of police station management. Taking advantage of this citizen interest, an alliance of domestic NGOs in 2006 organized the first Police Station Visitors Week: a mobilization of citizens—mostly in developing countries—who used a common protocol to visit police stations and score their performance on precisely these aspects of their operations and service to citizens. The annual event has been repeated regularly since then, the most recent at the time of writing taking place in October 2009, when more than 5,000 individuals visited and rated police

stations in 20 countries. The event is coordinated by the Altus Global Alliance with financial support from the UK Department for International Development.¹

The police stations are scored by the visitors, who each answer the same 20 questions immediately after the visit. The 20 questions consist of four questions in each of five areas of service: (i) community orientation; (ii) physical conditions; (iii) equal treatment of the public; (iv) transparency and accountability; and (v) detention conditions. Each question is answered on a 1-to-5 scale, where 1 is totally inadequate and 5 is excellent. The questions on transparency and accountability ask if there is information available to the public at the station about patterns of crime in the area, about the station's performance in terms of arrests made, crimes solved, or public satisfaction, about how to make a complaint against the police, and about the identities of each police officer or constable. The questions about detention conditions ask if the area is safely secured, if individual detainees are identified by name or numbers, if there are visiting facilities for lawyers, family members, or others, and if the conditions are sanitary.

The ratings are deliberately subjective. The point is not to make an objective assessment against some purported global norm, but instead to determine how well local citizens think that the station performs with respect to each issue area. At the same time, the ratings permit comparisons within a country, within a region, and globally.

The Police Station Visitors Week ratings are an example of participatory indicators where ordinary citizens not only join in the process of data collection, but control the standards embodied in the indicators. In each participating country, NGOs and other citizen organizations recruited, trained, and deployed citizen volunteers. Each NGO controls the recruitment in its area, and the recruitment strategies vary widely.

In Latin America, six countries participated in 2009: Bolivia, Brazil, Chile, Colombia, Mexico, and Peru. More than 1700 citizens visited 423 police stations in the region. The average overall scores were highest in Chile and Mexico, and lowest in Bolivia, but there was substantial variation from one station to the next in each country (see Figure 11.3). This variance drives competition among the stations. Public reports give special recognition to the highest scoring individual stations in each country, but police officials are also given private reports on the performance of each station, allowing national or state officials to pinpoint the weak stations depressing the national scores.

The Police Station Visitors Week scores do not make a perfect example of active indicators, for the scoring is only conducted annually, but in every other respect, the scores here meet the definition. They are calculated and presented to officials quickly, and they are responsive to changes in performance. Most important, the indicators are tightly linked to legitimate authority over the police officers whose

¹ In previous years, the Police Station Visitors Week has been supported by the Ford Foundation, the John D. and Catherine T. MacArthur Foundation, the Open Society Institute, and the Dutch Ministry of Foreign Affairs. The Police Station Visitors Week has continued since, most recently in October 2011.

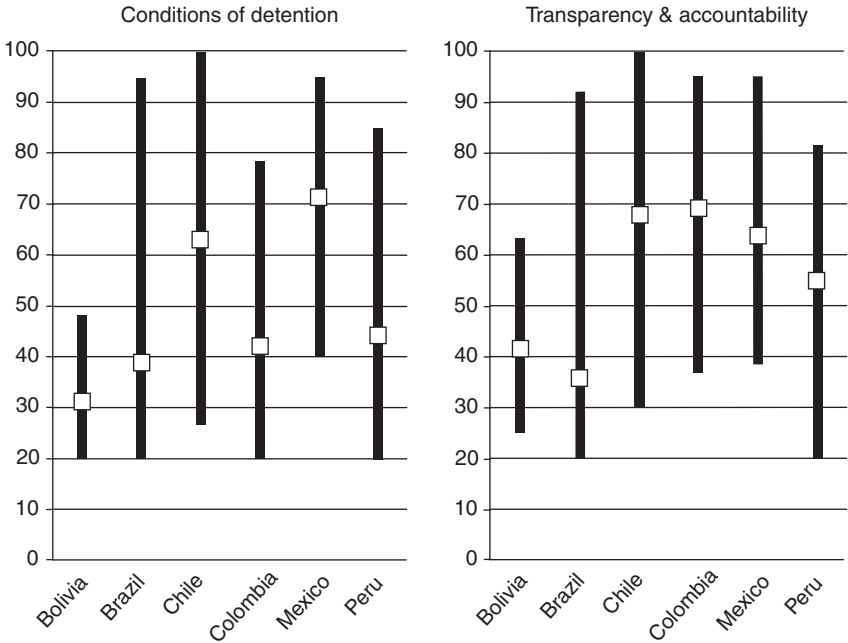


Fig. 11.3

Source: Altus Global Alliance, Police Station Visitors Week Global Report, 2009, Tables viii and ix, pp 70–1
Available at: <http://www.altus.org/pdf/apsvw_gr2009_en.pdf>.

work is being measured, but in this case the authority lies in the public, not in a senior official.

If the local NGOs who organize the visits in each city and country are successful at persuading the local news media to publicize the results or if the NGOs otherwise have the attention of police leadership, they can drive change through the indicators. The report on the 2009 round of Police Station Visitors Week contains several examples of just such reforms produced in earlier years, from the establishment of gender desks to the redesign of the physical stations themselves. The Royal Malaysian Police have gone so far as to include the 20 items that comprise the assessment into its own monitoring systems and it now invites citizens to visit and assess stations throughout the year, not only during the global Police Station Visitors Week—a further reminder that active indicators need to be refreshed more often than once a year.

Like the indicator of searches and hits in Jamaica, and the measures of pretrial detention in Lagos, Nigeria, the scores from the Police Station Visitors Week can usefully be aligned with longer-term national plans and the associated indicators. When a national planning agency or an international donor uses an active indicator as its indicator of progress, it leverages the power and authority of the user, whether a state official or a citizen. In either case, the authority of the longer-term program is enhanced by the incorporation of the active indicator and its sponsor’s power.

In contrast, when national plans and international donors insist on their own indicators of progress without regard to those being used by operational departments, they risk a power struggle.

The indicators generated by the Police Station Visitors Week are different in this respect from most other global indicators, in that they are not attempting to bolster the authority of their sponsor by reference to some purported international standard. The indicator is organized internationally, but the standards are local: how do local citizens understand the adequacy of each aspect of service being reviewed? In this way, they harness the authority of local citizens, rather than national standard-setting bodies.

With that contrast in mind, I turn now to the question of how these examples can help us design a system or framework of indicators that can be used across countries, in very different national contexts.

Global indicators from the bottom up

At a 2009 meeting at the Harvard Kennedy School, three officials from Sierra Leone listened as their counterparts from Nigeria reported on their efforts to develop an active indicator to manage the problems of pretrial detention. The idea of regularly collecting data from an exit sample appealed to them, and they hoped that they could emulate their West African neighbors, creating an indicator with which they could monitor the flow and duration of pretrial detention.

Similar opportunities for cross-national inspiration abound, for the priorities among leaders of the institutions of the justice sector in different countries are already surprisingly similar. The meeting at Harvard was simply one attempt to catalyze such inspiration, with teams from Jamaica, Nigeria, Papua New Guinea, and Sierra Leone each presenting the indicators they were designing to advance their own ambitions. The Sierra Leone team presented its effort to harmonize victimization survey data with administrative data on crimes reported to the police, to create a credible indicator of the level of crime. They left, however, determined to emulate the Nigerian indicator of pretrial detention.

In the months following that meeting, the construction of a pretrial detention indicator in Sierra Leone proved more difficult than the officials had expected. The exit data available in the Lagos prisons are not maintained at the remand prison in Freetown, so admission data became the only measure of flow, depriving the team of any data on length of stay. Rather than stop there, however, the team constructing the indicator turned to the courts, where—at the time of this writing—they have some hope of finding a different source of data on the duration of detention.

This suggests an iterative method for the construction of global indicators. Officials and technical experts could work together to identify local and domestic priorities, building active indicators for use by those with legitimate authority to drive improvements in performance. These examples could then be shared in any of several ways, but the sharing should be country-to-country, rather than through intermediaries, encouraging emulation. A further round of indicator development

would follow in each country, and then another opportunity to share the results and find inspiration for yet further work.

Dialogue is crucial in this iterative process, hence the need for direct country-to-country conversation. The dialogue is not merely technical, although it does inevitably turn to questions of data quality and availability, statistical calculation, and presentation. The dialogue is also substantive and normative, with participants debating what matters in the administration of justice and how chains of causality might operate. Do police reduce crime even when they seize a lot of guns or arrest many offenders? Which crimes matter most? How should we conceive of the role of private security when we assess the size and competence of a nation's police?

Yet, for all the uncertainties, the priorities that emerge from such dialogue suggest that a coherent set of global indicators could emerge relatively quickly. In just two years of such conversations among country teams at Harvard, the participants are already building consensus around five key aspects of their criminal justice systems needing indicators:

1. **Crime and Public Safety**, measured through a combination of survey data and officially recorded crime;
2. **Police Effectiveness**, measured in the immediate achievements of a variety of enforcement actions;
3. **Pretrial Detention**, measured by the number of long-term detainees whose cases are completed without further delay, alongside the minimization of detention times for minor offenders;
4. **Police-Prosecution Coordination**, measured by the speed and quality with which each responds to the other in the preparation of cases for trial;
5. **Coordination of the Formal and Customary Justice Systems**, measured by the frequency of recourse that each system has to the other.

This short list is not meant as a complete representation of any system, but it does map the actual concerns of officials from domestic governments and civil society committed to improving law enforcement and the administration of justice. More important, it suggests that a system of indicators built from the bottom up would not vary dramatically from a system built from the global level down. The difference lies in the opportunities for alignment, authorship, and linkage with legitimate authority.

PART IV

CASE STUDIES: ASSESSING THE STRENGTHS, PROBLEMS, AND EFFECTS OF INDICATORS IN HUMAN RIGHTS, HUMANITARIAN ASSISTANCE, AND SOCIAL INVESTMENT

This page intentionally left blank

12

Measuring Human Rights

*UN Indicators in Critical Perspective**

AnnJanette Rosga and Margaret L. Satterthwaite

Introduction¹

Debates over the best way to identify human rights violations, assess compliance with treaty obligations, and measure human rights progress over time have preoccupied scholars and practitioners for many years. Quantitative data has been forwarded as a central tool in the drive for better methods of assessment, monitoring, and advocacy. Among quantitative tools, human rights indicators have been identified as especially powerful. Rights indicators, “piece[s] of information used in measuring the extent to which a legal right is being fulfilled or enjoyed in a given situation,”² are understood to have a variety of advantages: they render complex data simple and easy to understand; they can be designed to demonstrate compliance with obligations, fulfillment of rights, and government efforts toward these goals; and they are capable of capturing progress over time and across countries. Since they are perceived to be an especially powerful intersection of law and social science, it is not surprising that NGOs, inter-governmental bodies, and governments have all begun to develop human rights indicators.

Human rights indicators are used to accomplish many, often contradictory, ends. They take their place among many manifestations of global governance projects. As such, they are situated at the nexus of international human rights

* This chapter is adapted, and draws significant portions of its text from: AnnJanette Rosga and Margaret L. Satterthwaite, “The Trust in Indicators: Measuring Human Rights,” *Berkeley Journal of International Law* 27 (2009): 253. The authors thank Valerie Brender (NYU J.D. 2012) for assistance in preparing the manuscript for this chapter. Work on this chapter was supported by the Filomen D’Agostino Research Fund at NYU School of Law.

¹ Each of the authors has been, in various capacities, involved in projects to help develop and/or analyze such indicators. As a cultural anthropologist (Rosga) and human rights legal scholar (Satterthwaite) respectively, we drafted the article from which this chapter is adapted as a way of thinking through the interdisciplinary functions of human rights indicators, as they seemed to embody an especially powerful intersection of law and social science.

² Maria Green, “What We Talk About When We Talk About Indicators: Current Approaches to Human Rights Measurement,” *Human Rights Quarterly* 23 (2001): 1062, 1065.

law, quantitative social science methodologies, administrative and regulatory apparatuses, advocacy projects, and the transnational spread of expert knowledges mobilized in the service of “standardization.”³

As early as the 1980s, assessments of the role of statistics in measuring human rights contained variously embedded discussions of indicators, including critical assessments of their use. A convergence of social, political, and economic forces and their accompanying epistemological shifts has dramatically increased demands for indicators without equal attention to their limitations. These demands arise not only from the perceived need within international human rights circles for better tools to hold governments to account, but also from the replication of verification and monitoring techniques used in a wide variety of business, non-profit, and governmental management contexts. In 1994, economic analyst Michael Power identified what he called an “audit explosion,” which he described as having “roots in a programmatic restructuring of organizational life and a new ‘rationality of governance.’”⁴ For Power, the audit, with its financial accounting origins, exemplifies both literally and metaphorically a number of monitoring and control practices characteristic of late modern social organization such as inspections, assessments, and other evaluative technologies.

Audit has become a benchmark for securing the legitimacy of organizational action in which auditable standards of performance have been created not merely to provide for substantive internal improvements to the quality of service but to make these improvements externally verifiable via acts of certification.⁵

Increasing demands for “indicators” are thus inextricable from the privileging of abstract, quantifiable, and putatively transferable data bits. As such, indicators partake of both the strengths and weaknesses of auditing practices. Without arguing whether indicators are inherently good or bad, this chapter suggests attention should be paid to how this growth is inextricable from an “accounting culture” in which tests of measurability prevail over accurate and contextually sensitive assessments of substance or actions.

We place efforts by United Nations (UN) bodies to create human rights indicators into conversation with scholarship on audit and standardization from the social sciences. While we are in agreement with the editors of this volume that there are very real drawbacks involved in any indicators project, we nevertheless conclude that debates about indicators may provide advocates with new opportunities to use the language of science and objectivity as a powerful tool to hold governments to account. Because human rights compliance indicators *can* threaten

³ See, e.g., Winton Higgins and Kristina Tamm Hallström, “Standardization, Globalization, and Rationalities of Government,” *Organization* 14 (2007): 685; Suzan Ilcan and Lynne Phillips, “Making Food Count: Expert Knowledge and Global Technologies of Government,” *Canadian Review of Sociology and Anthropology* 40 (2003): 441.

⁴ Michael Power, *The Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997), 10 (quoting Nikolas Rose and Peter Miller, “Political Power Beyond the State: Problematics of Government,” *British Journal of Sociology* 13 (1992): 173).

⁵ Ibid., 10–11.

to narrow the available space for democratic accountability and purport to turn an exercise of judgment into one of technical measurement, advocates of human rights should remain vigilant to the elisions at work in the indicators project. However, we argue that the conundrum of democratic accountability and the failure to clearly locate responsibility for judgment in international human rights assessment exercises are not themselves products of an over-reliance on quantitative, and thus acontextual, tools chosen to carry out those exercises. Rather, these are structural problems, foundational to international human rights law as it exists today. They would still be present even if quantitative indicators were banished from human rights assessment. Nevertheless, we join other contributors of this volume in describing the ways in which quantitative indicators lend themselves more easily to disguising issues of accountability and judgment as technical problems of measurement and data availability.

Indicating lack of trust: The evolving approach to human rights indicators

In a full-length article examining human rights indicators,⁶ we reviewed the history surrounding their use. Here, we offer only a brief summary of key points. On the whole, economic, cultural and social (ESC) rights as opposed to civil and political rights suffered a long-term marginalization, characterized by the late creation of a treaty-monitoring body for the International Covenant on Economic, Social, and Cultural Rights (ICESCR), and a lack of infrastructure for their advancement. In the early 1990s, human rights practitioners began searching in earnest for appropriate tools to turn the rhetoric of ESC rights into concrete reality. Indicators were among those tools, since they seemed to promise a way to monitor whether a state's conduct resulted in the fulfillment of rights.

In particular, indicators were perceived to be useful in measuring a state's progress over time and in helping to develop the core content of ESC rights. Indicators were also seen as allowing for comparison across countries. The Special Rapporteur and Committee on ESCR (CESCR) identified indicators as a way to make the seemingly vague obligations imposed on states parties by the ICESCR more concrete.

One of the moves the CESCR made to counter concerns about the vagueness of the ICESCR was to call on states to set up adequate means of monitoring their own progress in ensuring ESC rights. Here, the Committee reminded states that they must continually make good faith efforts to guarantee ESC rights for all, and that these efforts should be measurable.⁷

⁶ Rosga and Satterthwaite, "Trust in Indicators" (n.*), 253.

⁷ ECOSOC, CESCR, General Comment No. 3, "The Nature of States Parties' Obligations," (1990), reprinted in *Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies*, April 26, 2001, UN Doc. HRI/GEN/1/Rev.5, para. 10 at 11 (hereinafter "General Comment No. 3").

While states were not obliged by the treaty text to adopt any particular method for documenting and monitoring their progress in implementing Covenant rights, the Committee suggested that they should use benchmarks as “indication[s] of progress.”⁸ Importantly, these benchmarks were to be created and applied by the states, with the Committee in a supervisory, reviewing role. Finally—and perhaps most significantly—while resource constraints could legitimately have explained a state’s inability to fully implement each right for all individuals, they would not be allowed to excuse a failure to *monitor* state efforts toward full realization of ESC rights.

A few years after the CESCER made these recommendations, a UN seminar on “appropriate indicators to measure achievements in the progressive realization of economic, social and cultural rights” was held in preparation for the World Conference on Human Rights in Vienna in 1993. During the workshop, some key issues surfaced that will be explored below:

- (1) The problem of what was being “indicated” arose and appeared to be elided: what would be “indicated”—realization/enjoyment of rights, or compliance with the treaty?
- (2) The lack of clarity concerning the substantive content of the various rights under discussion: this was seen as a severe constraint in developing indicators.
- (3) Quantitative measures were seen to obscure “the qualitative and subjective nature of human rights.”⁹ However, the contours of the “subjective” nature of human rights were not discussed at length at the workshop.

The final relevant issue raised in the 1993 seminar was embodied in the conclusions of the conference: instead of producing a set of indicators to measure the core ESC rights, the conference concluded that it was impossible—at that early stage of the development of ESC rights—to identify and agree on indicators. Thus, the seminar ended with a non-conclusion: called together to agree on a set of indicators, the participants instead agreed that it was too early to identify appropriate indicators for rights whose contents remained indeterminate.

In the following years, the Committee continually requested that states parties develop and apply indicators to monitor their own progress in implementing various provisions of the treaty. The duty to monitor was also examined from the opposite side: in a General Comment, the CESCER asserted that a state’s failure to demonstratively monitor could *itself* amount to a violation of the Covenant.¹⁰

⁸ ECOSOC, CESCER, General Comment No. 1, “Reporting by States Parties” (1989), reprinted in *Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies* UN Doc. HRI/GEN/1/Rev.6 at 8 (2003), para. 6, at 14.

⁹ World Conference on Human Rights Report on the Seminar on Appropriate Indicators to Measure Achievements in the Progressive Realization of Economic, Social and Cultural Rights, April 20, 1993, UN Doc. A/CONF.157/PC/73, para. 108.

¹⁰ ECOSOC, CESCER, General Comment No. 14, “The Right to the Highest Attainable Standard of Health,” August 11, 2000, UN Doc. E/C.12/2000/4, para. 52.

This step taken by the CESCR—from suggesting to states that benchmarks might be “useful” in 1990, to asserting that the creation and use of monitoring systems including indicators is a treaty obligation from 1999 onward—is striking. In effect, it shifts the onus of conceptualizing and applying indicators from the international community to the states themselves. In relation to indicators, then, the Committee’s most vital role became the highly technical one of monitoring the state’s monitoring.

It is important to note, however, that the CESCR continued to express hope that universally applicable, rights-specific indicators could be developed. As will be explored below, the UN Office of the High Commissioner for Human Rights (OHCHR) aimed to make this hope concrete. Leaders of the various human rights treaty bodies—including the chairperson of the CESCR—requested that the OHCHR construct indicators for key human rights enshrined in the international human rights treaties. Since then, professional staff of the OHCHR, together with experts gathered from a variety of disciplines, have been hard at work constructing indicators to measure the efforts of states and the enjoyment of human rights all over the world.

Audit, distance, and the problem with trusting indicators

Leaving unresolved the question of whether to construct transnational or national indicators allowed the CESCR to hold in abeyance the difficult choice between either fully inhabiting the role of rights compliance monitors or completely embracing states parties’ control of the mechanics of measurement, thereby consigning itself primarily to the position of auditor. To understand how human rights indicators function as an *audit practice*, and further, to understand how audit practices bring human rights treaty bodies into the world of global governance, it will be useful to briefly detour away from legal scholarship into the social studies of science and technology.

Indicators as audit practice

Social scientists have noted that systems of auditing—in particular the language of quantification—are demanded when the following three conditions exist: one, there is “a relation of accountability” in which one party is mandated to provide an account of itself to another;¹¹ two, “the relation of accountability [is] complex such that [auditors] are distant from the actions of [auditees] and are unable personally to verify them”;¹² three, there are conditions of mutual distrust between the auditor and the auditee.¹³

¹¹ Power, *Audit Society* (n. 4), 5 (internal citation omitted).

¹² *Ibid.*, 5.

¹³ *Ibid.* Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1995) and Theodore M. Porter, “Objectivity as Standardization: The Rhetoric of Impersonality in Measurement, Statistics, and Cost–Benefit Analysis,” in Allan Megill (ed.), *Rethinking Objectivity* (Durham and London: Duke University Press, 1994), 197, 207.

In the field of human rights, all three conditions are met. First, states that have ratified the principal human rights conventions are required to provide to the various treaty bodies periodic accounts of their efforts to ensure those rights. Secondly, particularly in the international realm, distance between parties is created and maintained along numerous axes, including geography, language, culture, economic capacities, etc. Thirdly, distrust is common on the part of human rights monitors concerning governmental self-representation in the context of rights fulfillment and reporting. At the same time, states frequently mistrust those responsible for monitoring their human rights performance.

The best that the ESCR Committee could do under these conditions was to maintain a balance between mutually mistrusting parties—the auditor (in this case, the treaty body) and the auditee (in this case, the states parties). In ways that foreshadowed later tightrope-walking solutions arrived at by the OHCHR, the Committee effectively maintained this balance by turning to an audit-like structure. On one hand, a comprehensive set of international indicators created by the Committee might have been perceived as an imposition, suggesting the Committee's mistrust of states parties. This may have resulted in an exacerbation of any existing mistrust that states parties had of the Committee. In their favor, however, a set of international indicators could, by their very appearance of cross-national comparability, have offered the imprimatur of objectivity. If identical indicators were to be applied transnationally, no single state could argue it had been subjected to unreasonable, or unfairly targeted, standards of accountability.

On the other hand, a series of *nationally* constructed indicators, while allowing for important cultural and economic specificity, could by this very specificity risk appearing to both states and international treaty bodies as arbitrary. This is in part because international indicators are often assumed to be valuable only insofar as they are cross-nationally comparable.¹⁴ Nation-specific indicators might also risk exacerbating the Committee's mistrust of states parties, since work done to link rights with indicators might appear to provide opportunities for political manipulation.

The decision to abdicate the task of developing international indicators in favor of assigning itself the role of *monitor* of states' indicator efforts, handing over the task of indicator development to states themselves, was a rather neat solution, and a remarkable transformation for the Committee. For our purposes, it is most notable that the Committee undertook such an important shift in roles with so little discussion of its significance or ramifications.

The problems with trusting indicators

What are the ramifications of the Committee's removal from direct, substantive monitoring to what global governance scholars, to whom we shall turn below, have

¹⁴ See Hans-Otto Sano, "Human Rights Indicators: Purpose and Validity," Paper for Turku/Åbo Expert Meeting on Human Rights Indicators, March 11–13, 2005.

called “rule at a distance”? What significance can be read from the Committee’s effective abdication of one form of authority in favor of another in this instance? The significance lies in *audits* as a technology of control.

Audit practices entail “sampling, reliance on external expertise, and the assessment of internal control systems.” As Power argues,

[a]udits have value because they seek to draw general conclusions from a limited examination of the domain under investigation. But despite statistically credible foundations for sampling, audit practice is driven by economic pressures to derive more, or at least as much, assurance from fewer inputs . . . [R]eliance on others substitutes for directly checking the thing itself.¹⁵

Audits—and in our case, indicators—are further constrained by the limits of measurability and affordability. As former CESCR Chair Philip Alston explained, “For the most part, [indicators] are essentially statistical in nature. That in turn means that their subject matter must be potentially quantifiable, not only in a technical sense but in practical terms as well.”¹⁶ Chief among our concerns is the seemingly inevitable drift from this persistent demand for “potentially quantifiable” information to situations in which technical questions end up playing a more determinative role in the choice of human rights indicators than substantive considerations of the best way to assess rights. Questions such as, “Can it be counted? If so, when and how? How accurately? By whom?” are never merely technical. There are a number of important conceptual problems that beset measurement by indicator as well—problems that should be considered by human rights practitioners.

Numbers, statistics, and the language of quantification generally are seen as uniquely capable of reducing or eliminating subjectivity. In his seminal history of the association between objectivity and quantification, Theodore Porter observed that “quantification is a technology of distance”:¹⁷

The language of mathematics is highly structured and rule-bound. . . . In public and scientific uses . . . [it] has long been almost synonymous with rigor and universality. Since the rules for collecting and manipulating numbers are widely shared, they can easily be transported across oceans and continents and used to co-ordinate activities or settle disputes. Perhaps most crucially, reliance on numbers and quantitative manipulation minimizes the need for intimate knowledge and personal trust. Quantification is well suited for communication that goes beyond the boundaries of locality and community.¹⁸

Thus, the reliance on the language of quantification rests on an assumption that quantification will—at least partially—solve the problem of mistrust. The presentation

¹⁵ Ibid., 12.

¹⁶ Green, “What We Talk About When We Talk About Indicators” (n. 2), 1077 (quoting Philip Alston, “Concluding Observations,” in *Benchmarks for the Realization of Economic, Social and Cultural Rights: A Round-Table Discussion Organized by the High Commissioner for Human Rights*, Geneva, March 25, 1998).

¹⁷ Porter, *Trust in Numbers* (n. 13), ix.

¹⁸ Ibid.

of neatly tabulated numbers erases the means and messiness of their own generation. It obscures evidence of the human judgment involved in statistical production.

No one in the human rights field pretends that indicators can ever really be *apolitical*, but the need for information that is as accurate, reliable, and meaningful as possible is pressing. Unfortunately, discussions of criteria for good indicators tend not to specify which form of objectivity is at work. Wendy Lesser identifies two different senses of objectivity: the first sense of objectivity is the “sense that an objective report is disinterested, honest, reliable, impartial.”¹⁹ The second sense of the term suggests that “only something which is not subjective—which does not partake of the individual human viewpoint—can be fully objective, neutrally conveying things and events that are out in the world without the distorting coloration of human consciousness.”²⁰ Lesser points out that only a machine (her example is a television camera) can ever hope to approach the second sense:

And even that possibility seems remote . . . for in order to become a functional picture of reality, even television’s images need to be absorbed by our particular minds. The picture itself can have no meaning until viewers make something of it . . .²¹

But humans, with human judgment and interpretation, she reminds us, are necessary for the first sense.

I depend on people to give objective—in the sense of disinterested and impartial—interpretations to videotape Objectivity, in the first of the two senses, is a quality that only the human mind can have.²²

Discussions of “objective” indicators are vulnerable to two tendencies: first, to conflate these two senses of objectivity (with the result that ultimately those indicators requiring obvious human interpretation, such as qualitative assessments, are valued less highly), and secondly, to privilege those (generally numerical) indicators whose interpretive work is invisible.

The tendency for measures to become targets

As discussed with regard to the work of the CESC, chief among the strengths of auditing practices is their rhetorically powerful capacity for transferability. Indicators are said ideally to allow comparisons between nations at similar levels of economic development, and over time within a given nation.²³

¹⁹ Wendy Lesser, *Pictures at an Execution: An Inquiry into the Subject of Murder* (Cambridge: Harvard University Press, 1993), 139.

²⁰ Ibid.

²¹ Ibid.

²² As Power puts it, “the general principles of quality control systems . . . can be made to look similar and enable them to be compared at an abstract level.” Ibid. Exegeses and critiques of various conceptions of objectivity are numerous and reflect considerably more complexity and nuance than Lesser’s more succinct summation here. For an especially useful collection of essays, see *Rethinking Objectivity* (n. 13), 197.

²³ Power, *Audit Society* (n. 4), 12.

Yet even to the degree indicators “can be made to look similar and . . . compared at an abstract level,”²⁴ across geographical space, they tend to lose their efficacy as accurate and adequate measures over time. Scholars suggest that this is a characteristic of all measurement mechanisms that are tied to the goal of *improvement*. As the social anthropologist Marilyn Strathern puts it, “[w]hen a measure becomes a target, it ceases to be a good measure.”²⁵ Applied to the use of indicators in the human rights context, this principle explains the phenomenon of the “expectations gap”²⁶ in which a nation’s reporting of successful fulfillment of treaty obligations has a more or less distant relationship to the actual enjoyment of rights by its citizens.

[A]uditing works by virtue of actively *creating* the external organizational environment in which it operates. . . . Audit is never purely neutral in its operations. . . . New motivational structures emerge as auditees develop strategies to cope with being audited; it is important to be seen to comply with performance measurement systems while retaining as much autonomy as possible.²⁷

Applied to human rights indicators, this principle underlies the risk that, to the extent that governments do actively try to meet benchmarks and standards set in relation to international human rights treaties, the incentive to demonstrate success—or, say, “progressive realization”—according to given indicators may become greater than any incentive to substantively ensure the fulfillment and/or enjoyment of human rights themselves. For example, efforts abound to measure states’ compliance with the right to gender equality in education. A common indicator for this right is the ratio of girls to boys enrolled in primary education. Given that states will be rewarded for demonstrating narrow ratios, there is a built-in incentive to document female school enrollment. However, such figures do not allow substantive rights fulfillment to be assessed. Important contextual information that would do so includes the existence of curricula determined to be qualitatively equitable, the absence of sex segregation in schools, and actual school attendance of girls as compared to boys.

While the ratio of female to male enrollment may—when situated within sufficient contextual information—initially be a good indicator, the tendency for measures to become targets means that the link between the indicator and the right purportedly being measured attenuates over time. Thus, the demand for indicators to be “consistently measurable” carries with it an inherent weakness: applying the same indicators over time does not guarantee consistent measurement of rights fulfillment. Instead, indicators lose value as states adjust their practices to improve their standing according to those indicators.

²⁴ Ibid.

²⁵ Marilyn Strathern, “‘Improving Ratings’: Audit in the British University System,” *European Review* 5 (1997): 305, 308.

²⁶ Power, *Audit Society* (n. 4), 9–10.

²⁷ Ibid., 13.

Enter the experts: Renewed efforts to create international human rights indicators

As foreshadowed above, the CESCR did not abandon its hopes for universal indicators when it assigned itself the role of auditor. Instead, it turned to OHCHR experts for help. In so doing, they effectively asked a body of professional staff mandated to support the work of the treaty bodies to achieve what they themselves could not: to transform a judgment-laden process into one that appeared technical, scientific, and therefore—in a context in which the treaty bodies' authority is often in doubt—more legitimate.

Within a year, the OHCHR produced a report outlining a conceptual framework for indicators that seemed to assume an "appropriate" set of indicators would at once garner the support of social scientists, states, and civil society. The report can be read to suggest that such indicators could then be used by treaty bodies in an application of technical expertise, moving the treaty bodies beyond mere auditing to actual assessment of state compliance with the human rights standards set out in the treaties. Although not stated as such, this would seem to have two advantages. First, it would be an assessment that appeared to be objective because it was based on quantitative, scientifically validated methods, embodied in measurement indicators, rather than in more visibly subjective (and therefore more easily politicized) exercises of human judgment. Secondly, this very focus on indicators would effectively foreground the end-product of (apparently neutral) measurement made possible by indicators. This end-product would take the form of conclusions concerning states' progress on rights, compliance with treaty obligations, and recommended next steps. Simultaneously, the focus on indicators would *background* the acts of interpretation necessary to transform abstractly worded international laws into human rights standards both capable of and appropriate for transnational measurement.

The troubled authority of human rights treaty bodies in international law

This ambitious goal—to create a set of indicators capable of attracting agreement among states, human rights advocates, and social scientists—is understandable given the long-standing and unresolved issue of the status of the treaty bodies—and thus of their assessments—in international law. Indeed, the turn toward mechanics of measurement and notions of scientific objectivity may seem to offer a kind of authority that treaty bodies have never been able to achieve through the "quasi-judicial exercise[s]" that make up their core functions.²⁸

²⁸ OHCHR Report on Indicators for Monitoring Compliance with International Human Rights Instruments, May 11, 2006, UN Doc. HRI/MC/2006/7, para. 2 (hereinafter "2006 Report on Indicators").

While once the role of treaty bodies was understood to be almost entirely that of a supportive guide for states in implementing the treaties, their Committees now assess state performance through several procedures. All of the treaty bodies formally review state practice in hearings where state representatives are invited to present their periodic reports and to answer questions from treaty body members. These sessions are called “constructive dialogues,” and the official approach is non-adversarial. In practice, however, the “constructive” dialogues range from extremely collegial to contentious.

In recent years, a number of treaty bodies have taken significant steps to follow up on the recommendations that result from these hearings in the form of written concluding comments or observations. Informally, NGOs often use these in their advocacy efforts; consequently, these can become the subject of intense domestic and international pressure. The treaty bodies also issue General Comments. While General Comments began largely as vehicles to explain procedural matters or to provide guidance for states in preparing their reports to the committees, they have, over time, come to emphasize interpretation, explicating in some detail the substantive provisions of the relevant treaty. Some famous General Comments have sought to resolve—in favor of broad human rights principles—basic issues in international law; when they have done so, some states have strongly objected. Despite this, some General Comments have become extremely influential through formal and informal channels: guiding state policies, influencing UN agency actions, and becoming the framework for NGO action.

Treaty bodies that decide individual petitions have an even more judicial, or court-like, role than those that do not. Thus, the treaty bodies walk a difficult tightrope: constrained by positive international law, their greatest power is often normative. They are at the height of their authority when they are most persuasive, when their legal analysis—their *judgment*—is valued. A power based on persuasion can be severely limiting, however. In the case of indicators, as explored below, the treaty bodies seem to be hoping that the power of social science will have greater “compliance pull” than well-reasoned General Comments or persuasive decisions in individual cases.

Expert indicators: The OHCHR indicators initiative

To carry out the task entrusted to it by treaty bodies, the OHCHR itself turned to professionals, convening several meetings of experts from universities, international agencies, and NGOs, as well as members of the treaty bodies themselves, to discuss indicators. The result was a framework that made a significant contribution in terms of conceptually clarifying human rights compliance indicators, while also significantly scaling down expectations for the use of those indicators.

Whereas the initial ambitious hope was for a set of indicators that could be used for monitoring compliance with human rights treaties, the final product was a framework and attendant list of *illustrative* indicators that

allows a balance between the use of a core set of human rights indicators that may be universally relevant and at the same time retain[s] the flexibility of a more detailed and focused assessment on [sic] certain attributes of the relevant human rights, depending on the requirements of a particular situation.²⁹

In short, an effort that was initially aimed at giving the treaty bodies a new tool to help in the “quasi-judicial” exercise of assessing state compliance with treaties was transformed into an initiative aimed at giving all human rights practitioners a tool to conduct that assessment—implicitly now seen as a technical exercise.

From 2006 to 2008, the OHCHR piloted its framework and illustrative indicators through national and regional workshops.³⁰ Whereas earlier work seemed to assume that experts at the international level could develop universal indicators that would apply across countries, the 2008 report from the OHCHR calls for participation in the selection of indicators as an essential element of their use, though the means of such participation is not clearly specified. This participatory aspect is to be welcomed—indeed, as discussed below, it is essential. Still, the framework set out by the OHCHR reflects a continuing lack of clarity about a number of crucial issues. Perhaps most importantly, the OHCHR does not specify who, in addition to the treaty bodies, should *use* the indicators it has identified, instead suggesting they will be useful both for assessing compliance with human rights commitments, and for rights-based monitoring of development projects.³¹

Rights-based monitoring is an activity that is distinct from monitoring states’ compliance with human rights treaties. While it necessarily involves a close examination of states’ efforts in areas covered by relevant treaties, the goals of each type of monitoring are different. When assessing compliance with a treaty, the assessor is determining the extent to which a state has met its duties under a legal standard. When assessing a development project from a rights-based perspective, the assessor is determining the extent to which the project has advanced human development while also enhancing human rights. Indicators, therefore, are likely to differ significantly based on their use.

With respect to monitoring state compliance with human rights treaties, the OHCHR specifies that “[i]t is the objective of the work undertaken by OHCHR for the treaty bodies to identify relevant quantitative indicators that could be used in undertaking human rights assessments.”³² Toward this end, the OHCHR concludes that further work is needed to identify a “treaty-specific list of illustrative

²⁹ OHCHR Report on Indicators for Promoting and Monitoring the Implementation of Human Rights, June 6, 2008, UN Doc. HRI/MC/2008/3, para. 43 (hereinafter “2008 Report on Indicators”).

³⁰ See 2008 Report on Indicators (n. 29), paras 27–33.

³¹ The OHCHR gives general guidance only, stating that the “stakeholders who would be contributing to the monitoring process either as information providers, or as independent interpreters of the available information, or as the ultimate users of that information . . . may involve, inter alia, the national human rights institution (NHRI), the administrative agencies including the relevant line ministries as data providers, relevant non-governmental organizations engaged in monitoring human rights, consumer groups, other social groups, including parliamentary committees and claim-holders at large” (2008 Report on Indicators (n. 29), para. 37. See also paras 35–6.)

³² See *ibid.*, para. 35.

indicators.”³³ Given its determination that indicators also should be context-specific and participatory, the tension between state-specific and universal indicators appears to continue through the OHCHR’s indicators project. The specific roles that different actors (the treaty bodies, states, OHCHR, civil society) will have in selecting and using the various indicators in the monitoring process remain unclear, although OHCHR offers some suggestions to countries adopting its indicators framework on how that framework might be operationalized.³⁴

This is striking, not only since it represents a significant shift from the original task of identifying indicators for use by the treaty bodies, but also since there is enormous difference—legally and politically—among the various potential uses of the indicators forwarded by the OHCHR. For example, the treaty bodies have treaty-bestowed authority since they are charged with monitoring state compliance under the relevant treaty (though this authority is always contested); their use of the OHCHR indicators will carry with it a certain weight not present among other users. States have another type of authority—the type drawn on by the CESC when it has called on states themselves to create indicators; states’ adoption of the OHCHR indicators would go some distance toward legitimizing the framework, though it certainly would not have binding effect on other states under the international legal regime. The adoption and use of the OHCHR indicators by NGOs and other advocates would carry with it no special authority, though it would potentially lend an aura of legitimacy to the framework that may otherwise be lacking.

Finally, whether the indicators being designed actually do measure what they purport to measure is something that will need to be assessed over time. The OHCHR appears to hope that a core set of universal indicators can be agreed upon, but suggests that more contextual indicators will complement this core set. In its report, the OHCHR sets out “indicators for 12 human rights and the approach to the selection and contextualization of indicators with a view to encourage the application of the work at country level and in the treaty bodies.”³⁵ However, the OHCHR provides only general suggestions as to *who* may adapt the indicators at the national level, or what relationship such choices—when made by “monitoring stakeholders” other than the treaty bodies—will have on how treaty bodies use indicators when assessing state compliance with human rights law.³⁶

Instead of answering this question, the OHCHR presents the issue as a technical one, explaining that the framework presented “enables the potential users to make an informed choice on the type and level of indicator disaggregation that best reflects their contextual requirements for implementing human rights . . .”³⁷ Thus

³³ See *ibid.*, para. 7.

³⁴ Annex II to the 2008 Report on Indicators (n. 29), 34–50, calls for the compilation of “meta-data” to facilitate the identification and interpretation of sample indicators: e.g., definition, rationale, method of computation, sources, disaggregation, periodicity, comments, and limitations.

³⁵ *Ibid.*, para. 41.

³⁶ See 2008 Report on Indicators (n. 29), para. 37.

³⁷ *Ibid.*, para. 43. Rosga’s telephone interview with OHCHR indicators experts Nicolas Fasel and Grace Sanico Steffan in Geneva, Switzerland on November 15, 2010 provided information on how the

transformed, the issue of authority and judgment—always lurking behind the corner of the human rights regime—is again hidden from sight, buried in language concerning “informed choice” to be made by experts.³⁸

In this way, human rights indicators share the attributes of other types of standards. As Bengt Jacobsson has said, “[s]tandardization may be regarded as a way of regulating in a situation where there is no legal centre of authority.” Ominously characterizing the brave new “world of standards” that he and his colleagues set out to critically analyze, he continues:

[w]e will have a kind of symbolic and secularized society based on the premise that people voluntarily conform to the decisions of authorized expert knowledge. But while order is being established, responsibility may be vanishing.³⁹

Two responsibilities are at risk of vanishing in the context of human rights indicators: first, the responsibility for *transforming into measurable indicators* the more or less fully articulated normative standards that derive from international human rights treaties and the treaty bodies’ interpretations of them; and secondly, the *choice of indicators* that will be used to measure human rights commitments. The work of experts who designed these standards at the request of the treaty bodies effectively disappears in the final product: a neat set of one-page matrices that present structural, process, and outcome indicators for the major human rights set out in international human rights treaty law. Jacobsson points to three significant problems “related to standardization, which stem from reliance on experts: depoliticization, technicalization, and the emergence of regulation without responsibility.”⁴⁰

Office’s work has evolved since the 2008 Report. Following the multiple expert consultations, and sub-regional workshops with country-level stakeholders in Asia, Africa, and Latin America, the OHCHR organized, at the request of national stakeholders, workshops and consultations involving national human rights institutions (e.g., Human Rights Commissions accredited by the United Nations), government and statistical agencies, civil society organizations, and UN country teams. “This work is not done only in the context of reporting to Treaty Bodies, but also in relation to national human rights action plans and for mainstreaming human rights in development plans . . . The emphasis is on national relevance and not transnational comparability.” OHCHR published an update on its indicators project in 2011, pointing out that the treaty bodies had endorsed the OHCHR indicators framework and the CESCR had referred to it in its 2008 reporting guidelines (paras 1, 41); noting that treaty bodies, the Human Rights Council, and UN Special Rapporteurs are increasingly using some form of indicators (paras 41–7); and calling for greater collaboration among government agencies, human rights organizations, and statistical bodies (para. 51). The 2011 Report places the OHCHR indicators project in the context of a broader turn toward quantitative analysis and “evidence-based” human rights monitoring (para. 53). OHCHR Report of the United Nations High Commissioner for Human Rights, April 26, 2011, UN Doc. E/2011/90 (hereinafter “2011 Report on Indicators”).

³⁸ It is important to note that we are not attributing any intentionality to the OHCHR in the generation of these effects; that is, we do not argue that the OHCHR’s *intent* has been to obscure the role of judgment through the use of indicators. Rather, we are calling attention to the ways in which this elision is an effect of the increasing reliance on social science experts, and an especially pronounced effect when quantitative tools predominate. In its 2011 Report on Indicators, the OHCHR includes a section emphasizing that “[t]he use of indicators does not replace the normative analysis of a human rights situation.” 2011 Report on Indicators (n. 37), para. 18.

³⁹ Bengt Jacobsson, “Standardization and Expert Knowledge,” in Nils Brunsson et al. (eds), *A World of Standards* (Oxford and New York: Oxford University Press, 2000), 40.

⁴⁰ Jacobsson, “Standardization and Expert Knowledge” (n. 39), 49.

Why is this a problem? Jacobsson argues that the danger lies in the potential for standards to depoliticize choices otherwise openly contested in the public sphere. Here, the marketing by OHCHR of its indicators seems intended to bring states *voluntarily* into closer line by persuading those with whom they work—"human rights stakeholders"—that indicators offer a technical answer to what would otherwise appear to be judgment-laden (and thus court-like) or highly contested (and thus political) issues. The auditing role of the CESCR is thereby maintained, but now with states being asked to use the rules set out by international experts rather than those of the state's choosing. In this way, what might otherwise appear as a bold assertion of authority and power by the treaty bodies is passed off as a technical exercise that should be voluntarily accepted by rational human rights practitioners—including those working for the state.

In the end, however, this effort cannot solve the problem that generated the CESCR's audit practice to begin with: the relationship of distrust between the treaty bodies and the states whose efforts they monitor. This is because, although it appears to do so, the framework forwarded by the OHCHR will never be able to do the real work of assessing where states have fallen short of their obligations under human rights treaties. To take one example: imagine an assessor seeks to determine the adequacy of a state's allocations to primary education and to the promotion of higher education for women. Imagine further that the state in question has an extensive primary school system, but that it has systematically undervalued women's roles in the professions. How will the user of the indicator "share of public expenditure on education devoted to primary education" know whether a state's choice to allocate proportionally less money to primary education than it allocates to scholarships and professional training for women is permissible when simultaneously confronted with the indicator "proportion of females with professional or university qualification"?⁴¹

Such can never be a technocratic assessment. It requires, instead, the exercise of *judgment*. By evading possibly the most thorny issue—who will be the final arbiter of which indicators will be used, and how exactly they will be used to assess state compliance with international human rights law—the OHCHR evades one of the most difficult issues in human rights law: that of authority. In the end, the discussion of human rights indicators requires us to attend to the issue of judgment, and the unique challenges posed by a system of law that fails to locate authority for judgment in any given body. In implicitly recognizing this problem but apparently seeking to elide it, the OHCHR deploys the language of expertise. While its framework for human rights indicators is conceptually clear and may allow for powerful advocacy, it does not resolve the underlying problem that its evident trust in numbers seeks to fix—the pesky, irreducible core of human judgment.

⁴¹ Both of these indicators are included as illustrative of the right to education. See 2008 Report on Indicators (n. 29), para. 28.

Human rights indicators as technologies of global governance

In many ways, the turn to indicators in the human rights context mirrors trends across the landscape of transnational governance. These trends, manifested in part by audit practices, also come under critique from across the disciplines under the rubric of “global governance” and “governmentality” analyses. The gist of these critiques is well summarized by Jacobsson, who warns that:

[g]reater reliance on standards may involve a danger that so-called technical expert knowledge will become a substitute for ethical and political discourse . . . There will be a growing focus on how things are done—a focus on form rather than content.⁴²

The move to audit by the CESCER may embody this shift from content to form. Further, it risks displacing contestation over substantive rights issues onto seemingly bureaucratic or technical decisions about choice of indicators. This is—in part—the promise and peril of numbers. In her ethnographic project studying rights indicators, Sally Engle Merry lodges a similar critique:

Numbers are the epitome of the modern fact because they seem to be simple descriptors of phenomena and to resist the biases of conjecture and theory since they are subject to the invariable rules of mathematics. Numbers have become the bedrock of systematic knowledge because they seem free of interpretation, as neutral and descriptive. They are presented as objective, with an interpretive narrative attached to them by which they are given meaning.⁴³

As the treaty bodies turned to the OHCHR for assistance with the development of universal human rights indicators, thereby extending the turn to expert assistance, it would seem that this danger has, if anything, increased. Scholars who study globalization practices have noted the many ways in which the turn to technocratic numeracy can result in bureaucratic stalemates at best and a range of negative unintended consequences at worst.⁴⁴

One of the unintended consequences may be the tendency of compliance indicators to close down spaces for democratic contestation. In carrying out its human rights obligations, states must continually make difficult prioritization choices, especially in the context of economic and social rights. Which rights should receive the state’s most concentrated attention—the right to potable drinking water or the ability to access childhood vaccinations? Gender equality in higher education or ensuring non-discriminatory employment conditions? How should government authorities *make* such determinations? Who should decide and by

⁴² Jacobsson, “Standardization and Expert Knowledge” (n. 39), 46.

⁴³ Sally Engle Merry, “Measuring the World: Indicators, Human Rights, and Global Governance,” *Current Anthropology* 52 (2011): S83–95 (citing Mary Poovey, *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society* (Chicago: University of Chicago Press, 1998)).

⁴⁴ See, e.g., Angela P. Cheater, “Globalisation and the New Technologies of Knowing: Anthropological Calculus or Chaos?” in Marilyn Strathern (ed.), *Shifting Contexts: Transformations in Anthropological Knowledge* (London and New York: Routledge, 1995), 117–30.

what means? And most relevant to our purposes, what does human rights law have to say about these challenging questions, and how might indicators measure state efforts to answer them?

In many respects, human rights law has very little to say of a concrete nature in answer to these questions. Prioritization challenges are often answered with the demand that all rights receive equal attention and priority—a claim often made by those invoking the principle of indivisibility.⁴⁵ In practice this is seldom, if ever, possible at the level of everyday policy—especially in resource-poor countries. Moreover, this principle does not mean very much in practice, since governments formulating policy are continually forced to stress some policies or actions over others. However, while there may not be clear *answers* to the question of prioritization, a good deal of work has been done to establish more concrete standards to *guide governments' choices* concerning competing rights demands.

The concept of minimum core obligations requires that states act to immediately fulfill certain basic rights standards, regardless of the economic status of the country, thus forcing prioritization of actions to achieve those goals. The principles of non-discrimination and equality have been interpreted to impose duties on states to ensure that they immediately prioritize the rights of the most vulnerable and marginalized communities. Moreover, the principle of non-retrogression means that governments must ensure that their policies and actions are designed to ensure that rights fulfillment is not diminished, but instead progresses forward toward full enjoyment.

These principles provide a good deal of guidance to states seeking to uphold their human rights obligations, and in some scenarios, states may rely upon them to prioritize activities. However, the principles will often fail to provide the answers to questions about prioritization and emphasis in implementation. Only rarely will they provide a rule of decision for policymakers choosing among options for actions that can improve human rights. This gap—between international norm and domestic implementation—is both normal and desirable. It opens space for states—and more importantly, for national populations within states—to determine how best to carry out their duties. Within this space, democratic contestation and participation by those most directly affected can take place. Without such a gap, human rights law would perversely short-cut democratic processes by imposing specific policy choices on states.

Indicators may threaten to close this fruitful gap between international law and domestic policy by, for example, targeting through measurement the outcomes of certain policies, or even turning specific policy choices themselves into indicators. For an example of the latter, the indicator “coverage of targeted population covered under public programs on nutrition education and awareness” has been identified by the OHCHR as an indicator of the right to food.⁴⁶ While such programs are one way to achieve an important element of the right to food, they might not be the one

⁴⁵ See, e.g., World Conference on Human Rights Vienna Declaration and Programme of Action, July 12, 1993, UN Doc. A/CONF.157/23, para. 5.

⁴⁶ 2008 Report on Indicator (n. 29), para. 24.

preferred by the inhabitants of a specific state, who may want funds to be spent on direct food aid or supplements to farmers who cultivate staple foods. The OHCHR's recent work within particular countries, however, seems to prioritize the "right to participation" by involving a variety of local actors, including national human rights institutions, government agencies, statistical offices, and civil society organizations. According to OHCHR's indicators experts,

At the country level, [actors from civil society, government, and national level human rights institutions] start the development of indicators, either with one or two, or several sets of rights, or they try to use/incorporate the OHCHR conceptual and methodological framework, and [illustrative] list as well, into National Action Plans. . . . We try to develop a dialogue between the various agencies, many of whom are often sitting down together for the first time.⁴⁷

Further, given the problems of measurability and availability of data, there may be a tendency to choose indicators that capture the outcomes of the most easily—or the most consistently—measurable policies or programs. For example, "proportion of targeted population covered under public nutrition supplement programs" has been used by OHCHR as an indicator of the right to food. This indicator asks the government to count something it is very likely already counting—the number of households benefitting from its own ("public") nutrition supplement programs. This indicator may reveal a great deal in countries that have chosen to respond to hunger and malnutrition with state-run supplement programs. It will be less revealing of the extent to which the right to food is being fulfilled in countries where the non-profit sector has taken on greater food supplementation duties than the state, however. Because of the bias in indicator construction toward choosing easily countable phenomena, one kind of indicator is more likely to be chosen than another. Finally, indicators that measure the outcome of only certain specific policies, and indicators that mandate specific policy choices, have an attenuated relationship to the legal norm at issue. They are therefore ill-suited for use in monitoring compliance with legal duties.

In sum, the development and use of compliance indicators may have the tendency to artificially close the gap between international law and domestic policy, thus shrinking the required spaces for participation. If indicators are designed and imposed uniformly across countries, and if they are not susceptible to being calibrated according to national priorities and deliberation, they could backfire as accountability mechanisms.

So it may seem that there is little to applaud in this recent manifestation of the "turn to indicators"—however tentative and "illustrative," however strewn with caveats they might be—by the OHCHR. And yet, as we have traced the human rights treaty bodies' ongoing efforts to grapple with the task of holding states accountable to their commitments to human rights treaties, we have come to

⁴⁷ Interview with Fasel and Sanico Steffan, November 15, 2010. One example of public consultations being carried out for the process of developing national (British) indicators can be found at: <<http://personal.lse.ac.uk/prechr/>>.

appreciate new aspects of this project that a “governmentality”-focused analysis risks occluding.

Such an analysis might focus solely on the dangers inherent in the turn to indicators—on the evasion of difficult questions of judgment represented by the human rights community’s embrace of technocratic numeracy, and on the concomitant submersion of political debates “by technical questions of measurement, criteria, and data accessibility.”⁴⁸ It would be easy enough to apply a similar analysis to the work of treaty bodies and the OHCHR in developing international human rights indicators. Certainly the turn to experts for putatively independent, bias-free, and scientifically valid techniques with which to assess the degree to which states are living up to the commitments they make represents the OHCHR’s participation in wider transnational governance trends.

Through human rights indicators, the CESC and OHCHR are certainly promoting engagement with various international standards in order to shape the conduct of governments—that is, they are intending to “transform the terrain of government policies.”⁴⁹ However, there are significant differences that should be noted as well. Much of the governmentality literature, while helpfully diversifying our conceptions of those who *govern*,⁵⁰ nonetheless often implicitly assumes—through its choice of examples—that the targets of governance are largely if not solely citizens and populations. Human rights indicators, on the other hand, emerge out of projects aimed at changing the conduct of *governments* toward those same populations.

We believe that this difference matters—that the power of indicators, when harnessed by human rights advocates, may be fruitfully turned on the state by those the state has failed to serve, or even harmed.⁵¹ Indeed, we believe that human rights indicators—if designed with these valences of power in mind—can be used to monitor whether governments have arrived at effective human rights policies and actions through democratic processes. In other words, instead of disappearing politics, indicators *should* be designed to allow for the monitoring of governmental processes to ensure they are participatory and open to deliberation and debate.

Discussions of indicators need not be technical conversations devoid of political contestation. Nor must they be conversations in which participants are seeking to submerge difficult questions of judgment in the abstract language of numbers.

⁴⁸ Merry, “Measuring the World” (n. 43).

⁴⁹ Suzan Ilcan and Lynne Phillips, “Making Food Count: Expert Knowledge and Global Technologies of Government,” *Canadian Review of Sociology and Anthropology* 40 (2003): 441, 444.

⁵⁰ See Andrew Barry, “Ethical Capitalism,” in Wendy Larner and William Walters (eds), *Global Governmentality: Governing International Spaces* (London and New York: Routledge, 2004), 195, 202 (“Writers on governmentality, following Foucault, have long emphasized that the activity of government cannot be reduced to the actions of the state. In an era where direct state control and ownership has declined . . . international institutions, NGOs, auditors, consultants and multinational corporations are together expected to perform the job of government at a distance”).

⁵¹ For an example of recent scholarship making this argument, see Sital Kalantry, Jocelyn E. Getgen, and Steven Arigg Koh, “Enhancing Enforcement of Economic, Social, and Cultural Rights Using Indicators: A Focus on the Right to Education in the ICESCR,” *Human Rights Quarterly* 32 (2010): 254.

They should be conversations in which engaged social actors are grappling with the very phenomena we have been describing—actors who are fully aware of both the power and the limits that statistics possess.

Conclusion

The value of indicators as a social technology can neither be determined in advance, nor assessed on the basis that they draw on the power of quantitative language. While it may be true that quantitative methods, in their very abstraction and stripping away of contextualizing information, pose particular—and especially high—risks of misuse by those with the power to mobilize them, they are tools like any other. All tools can be misused; all social actors with power can misuse that power. The key lies in knowing where—and how—human judgment and political contestation should enter. Rather than trusting in numbers too quickly, those using human rights compliance indicators should embrace the opportunities presented by this new project, finding ways to utilize human rights indicators as a tool of global governance that allow the *governed* to form strategic political alliances with global bodies in the task of holding their governors to account.

The Use of Indicators to Measure Government Responses to Human Trafficking

Anne T. Gallagher and Janie Chuang*

While its precise definition is complex and contested, human trafficking is generally understood to refer to the movement or maintenance of individuals in situations of exploitation through—in the case of adult victims—force, deception, coercion, or other means. Contemporary forms of trafficking, which feed cheap or free labor into sectors as diverse as agriculture, fisheries, construction, prostitution, and domestic work, are a feature and product of our globalized economy. Those who profit from exploiting the labor of others are reaping the benefits of inequalities within and between countries; the age-old human compulsion to move in search of a better life; and, in the case of cross-border trafficking, rapidly diminishing opportunities for safe, legal, and gainful migration.

Although various aspects of trafficking have been subject to international legal regulation since the early 1900s, it is only over the past decade that a detailed legal regime has emerged around this issue. The centerpiece of that regime is a protocol to the United Nations Convention against Transnational Crime, adopted in 2000 (Trafficking Protocol).¹ In addition to articulating the first-ever international legal definition of trafficking,² the Protocol sets out, in considerable detail, the steps to

* Note that the views expressed in this chapter should not be taken to reflect those of the organizations with whom the author is or has been associated.

¹ Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, Supplementing the UN Convention against Transnational Organized Crime, GA Res. 55/25, Annex II, November 15, 2005, UN Doc. A/55/383 (hereinafter Trafficking Protocol).

² Article 3 of the Trafficking Protocol defines trafficking as: “(a) . . . the recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, or deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labour or services, slavery or practices similar to slavery, servitude or the removal of organs; (b) the consent of a victim of trafficking to the intended exploitation set forth in subparagraph (a) shall be irrelevant where any of the means set forth in subparagraph (a) have been used.” The Protocol subsequently notes that the “means” element is not required in cases where the individual involved is a child. For a detailed discussion of the definition, including of ongoing controversies around certain of its elements see chapter 1 of Anne T. Gallagher, *The International Law of Human Trafficking* (Cambridge: Cambridge University Press, 2010).

be taken by states parties in preventing and dealing with this crime. Since its adoption, the Protocol has been supplemented by a multitude of international and regional instruments that, with only a few exceptions, have added considerably to our understanding of rights and obligations in this area. One particularly significant example is the European Trafficking Convention, adopted by the Council of Europe in 2002.³

An equally spectacular shift has taken place at the national level. Trafficking is now criminalized in almost every country, with most taking at least some steps to also provide protection and support to victims.⁴ In many cases, statutory reforms have been accompanied by a strengthening of criminal justice institutions and procedures aimed at ending the high levels of impunity traditionally enjoyed by those who profit from the exploitation of others and, at least in some instances, securing justice for victims.⁵ The United States has gone even further, using its domestic legislation (the Trafficking Victims Protection Act of 2000 (TVPA))⁶ to take the war against trafficking beyond its borders by way of an oversight mechanism that now reaches almost every country on the planet.

In addition to nurturing great legal and normative change, global concern about trafficking has also prompted massive investment by inter-governmental organizations, states, and civil society into anti-trafficking interventions. This in turn has fed a demand for tools and mechanisms that will help make sense of the problem and validate what is being done: by, for example, measuring the true extent of trafficking or evaluating the absolute and comparative worth of an individual state's response. A rapid rise in the formulation and application of "indicators" is one manifestation of the new environment within which trafficking is being discussed and addressed. In this area, as in many others, the development of indicators represents an attempt to impose order over chaos; to simplify the complex; to measure the potentially immeasurable; and to reduce the possibly irreducible. While these efforts are understandable, it is essential to acknowledge that certain features of trafficking serve to compromise the utility and feasibility of an indicators-based approach—particularly in relation to measuring government responses. Trafficking is a complex issue and the current knowledge base is very weak. There is vigorous disagreement as to how trafficking should be understood and responded

³ See, e.g., Council of Europe Convention on Action against Trafficking in Persons, Council of Europe Treaty Series, No. 197, 16V.2005 (hereinafter European Trafficking Convention).

⁴ UN Office on Drugs and Crime (UNODC), Global Report on Trafficking in Persons, February 2009, 22 (hereinafter UNODC Global Report). US Department of State, Trafficking in Persons Report 2010 (hereinafter 2010 TIP Report).

⁵ See generally Anne Gallagher and Paul Holmes, "Developing an Effective Criminal Justice Response to Human Trafficking: Lessons from the Front Line," *International Criminal Justice Review* 18 (2008): 318 (hereinafter: Gallagher and Holmes); and UNODC Global Report (n. 4).

⁶ Trafficking Victims Protection Act of 2000, Pub.L.No. 106–386 § 2A, 114 Stat. 1464 (2000) (hereinafter TVPA), as supplemented by the Trafficking Victims Protection Reauthorization Act of 2003, Pub. L.No. 108–193, 117 Stat. 2875 (hereinafter 2003 TVPRA), the Trafficking Victims Protection Reauthorization Act of 2005, Pub.L.No. 109–164, 119 Stat. 3558 (2006) (hereinafter 2005 TVPRA), and the William Wilberforce Trafficking Victims Protection Reauthorization Act of 2008, Pub.L.No. 110–457, 122 Stat. 5044 (2008) (hereinafter 2008 TVPRA) (codified at 22 USC §7101).

to. Perhaps most importantly, the relevant regulatory framework around this issue is young—and the fundamental assumptions on which it is based are largely untested.

This chapter considers these issues within a broader assessment of how indicators function in and affect the field of human trafficking. The first Part examines the various ways in which the language and idea of indicators have been used in the trafficking context to: (1) track *indications* of the trafficking phenomenon; (2) measure the *performance* of specific anti-trafficking interventions; and (3) assess national *compliance* with a particular legal or policy regime. The second part provides an in-depth examination of indicator use in the third area through a case study. The subject of this analysis is the prominent and controversial indicator-based compliance mechanism established by the US State Department. This Part begins by examining the development of the US TIP Indicators, specifically, the rationale behind the indicators' creation, and the indicators' evolution over time. It then explores their application, including the methodology for data collection and analysis and the ways in which indicators are weighed, interpreted, and expanded in practice. The final Part concludes with a consideration of the legal, normative, and policy implications of the US TIP Indicators, exploring their impact on target states as well as on international law and policy.

Overview of trafficking-related indicators

There is little agreement on the function and application of indicators amongst those working and writing in the field of human trafficking. This Part provides an overview of the various ways in which the language and idea of indicators have been used in this sector, drawing on contemporary examples, one of which is subject to much more detailed examination in the next Part.

Indicators as indications

In trafficking-related literature and research, the term “indicators” is most commonly used to refer to *indications* of trafficking, i.e., those signs that will enable a determination to be made that an individual—perhaps an unaccompanied child, a “rescued” sex worker, or a smuggled migrant—has been trafficked.⁷ Less typically, indicators are formulated to predict trafficking: either within a given community or social group or from or to a particular country.⁸ The importance attached to indicator/indications confirms that the emergence of an internationally agreed

⁷ Frank Laczko and Marco Gramegna, “Developing Better Indicators of Human Trafficking,” *Brown Journal of World Affairs* 10 (2003): 179.

⁸ Kevin Bales, “What Predicts Human Trafficking?,” *International Journal of Comparative and Applied Criminal Justice* 31 (2007): 269; Andrea Cirineo and Sacco Studnicka, “Corruption and Human Trafficking in Brazil: Findings from a Multi-Modal Approach,” *European Journal of Criminology* 7 (2010): 29.

definition of trafficking has not served to eliminate uncertainties in relation to identification of trafficking cases, of victims, and of perpetrators.

The conception of indicators as indications is most clearly exemplified in the United States *Trafficking in Persons Reports* [TIP Report]. The 2011 TIP Report, for example, advises governments to proactively investigate “trafficking indicators,”⁹ particularly when dealing with undocumented migrant workers¹⁰ and prostitutes;¹¹ and notes with approval the development of mental health indicators to facilitate victim identification.¹² It is only with respect to Estonia’s efforts to develop indicators for evaluating the success of anti-trafficking programs¹³ that indicators are used in reference to assessing *responses* to trafficking. In similar fashion, the International Labour Office has undertaken extensive work on “indicators of forced labour”: those signs or features of a particular situation that help to separate out difficult or even exploitative labor from that which is truly “forced.”¹⁴ Building on this work, the ILO and the European Commission subsequently brought together experts to develop consensus on “operational indicators of trafficking in human beings.”¹⁵ According to the sponsoring organizations, such indicators were necessary given the lack of clarity around certain aspects of the definition of trafficking such as “coercion” and “consent.” They argue that without operational indicators, there is: “a risk that researchers and practitioners may not recognize trafficking when they see it—or see trafficking where it does not exist.”¹⁶ Anti-trafficking training materials and resources of the United Nations Office on Drugs and Crime (UNODC) include lists of both general and situation-specific “indicators of trafficking.”¹⁷ Major international and regional policy documents on trafficking have used the term in the same way.¹⁸

Indicators of performance: anti-trafficking interventions

The anti-trafficking sector is relatively new, very well funded, and consequently crowded. Governments, inter-governmental organizations, and NGOs are involved, variously, as donors, recipients, and implementers. As typically happens in such

⁹ US Department of State, *Trafficking in Persons Report 2011*, 40 (generally), 313 (Senegal), 146 (Djibouti) (hereinafter 2011 TIP Report).

¹⁰ 2011 TIP Report (n. 9), 206 (Japan), 384 (Vietnam).

¹¹ 2011 TIP Report (n. 9), 236 (Luxembourg), 322 (Singapore).

¹² 2011 TIP Report (n. 9), 303 (Qatar).

¹³ 2011 TIP Report (n. 9), 157 (Estonia).

¹⁴ International Labour Office (ILO), “Summary Report: Technical consultation on forced labour indicators, data collection and national estimates” (2006) available at <http://www.ilo.org/wcmsp5/groups/public/-ed_norm/-declaration/documents/meetingdocument/wcms_081981.pdf>.

¹⁵ International Labour Office (ILO), *Operational Indicators of Trafficking in Human Beings*, 2009, available at <http://www.ilo.org/sapfl/Informationresources/Factsheetsandbrochures/lang-en/docName-WCMS_105023/index.htm>.

¹⁶ ILO, *Operational Indicators of Trafficking in Human Beings* (n. 15).

¹⁷ See, e.g., UN Office on Drugs and Crime, *Toolkit to Combat Trafficking in Persons* (2008); UN Office for Drugs and Crime, *Anti-Human Trafficking Manual for Criminal Justice Practitioners* (2009).

¹⁸ See, e.g., Brussels Declaration on Preventing and Combating Trafficking in Human Beings (2002).

situations, initial waves of intervention took place in a performance evaluation vacuum. This was noted by, amongst others, the US Government Accountability Office which, in 2006, criticized the absence of “measurable goals and associated indicators to evaluate the overall effectiveness of [US] efforts to combat trafficking abroad,”¹⁹ and, echoing a recommendation made by the State Department’s Inspector-General, called for “performance indicators to compare progress in combating trafficking from year to year.”²⁰ Time, experience, and improved understandings have all contributed to heightened expectations, particularly on the part of donors. The anti-trafficking sector is now, at least in theory, part of the audit culture: it is required to meet those standards that have been established generally for development cooperation with regard to confirmation and documentation of impact, effectiveness, and value for money. Increasingly, this is leading to calls for formulation of trafficking-specific performance indicators: measurable characteristics or attributes that can be used to assess a particular activity or intervention or to ascertain progress towards a specified goal. The quality of responses to such calls has varied widely but a cursory review of indicators attached to major anti-trafficking programs and projects reveals a worrying lack of sophistication and nuance. There is, for example, little common agreement about what constitutes success and a general reluctance to acknowledge obstacles to securing data. A strong preference for measuring simple outputs rather than true impact or effectiveness is a logical consequence of this situation.²¹

There has been some modest progress in developing indicators of potentially greater sophistication and broader applicability. The International Organization for Migration (IOM), for example, recently released a handbook on performance indicators (the *Handbook*) for managers, developers, implementers, evaluators, and donors working on counter-trafficking projects.²² Developed in close collaboration with the US State Department, the *Handbook* seeks to move beyond measuring the immediate results of interventions (outputs), and towards the creation of “a method of global comparability to help understand the overall impact [of individual projects].”²³ This approach follows the accepted hierarchy of development-related monitoring and evaluation in which impact assessment (as opposed to output assessment) is seen as the most critical. For IOM, indicators can be either

¹⁹ United States Government Accountability Office, “Human Trafficking: Better Data, Strategy and Reporting Needed to Enhance UN Antitrafficking Efforts Abroad” (2006), 24 (hereinafter GAO, Human Trafficking).

²⁰ GAO, Human Trafficking (n. 19), 25.

²¹ See further, Global Alliance against Trafficking in Women, “Feeling Good about Feeling Bad . . . A Global Review of Evaluation in Anti-Trafficking Initiatives” (2010), available at <http://www.gaatw.org/publications/GAATW_Global_Review.FeelingGood.AboutFeelingBad.pdf>.

For a detailed consideration of these issues in the specific context of criminal justice interventions, see Anne T. Gallagher and Rebecca Surtees, “Measuring the Impact and Effectiveness of Anti-Trafficking Interventions in the Criminal Justice Sector: Who Decides, and How?” 1 *Anti-Trafficking Review* (2012) (forthcoming) (hereinafter Gallagher and Surtees, “Measuring the Impact and Effectiveness of Anti-Trafficking Interventions in the Criminal Justice Sector”).

²² International Organization for Migration (IOM), *Handbook on Performance Indicators for Counter-Trafficking Projects* (IOM, 2008).

²³ IOM, *Handbook on Performance Indicators* (n. 22), 8.

qualitative or quantitative, used to measure, monitor, or evaluate performance as well as to assess change. Clearly there is much more work to be done. A 2011 evaluation of IOM's anti-trafficking program criticized the Organization for not following its own policies with respect to indicators and further noted the weakness of the IOM indicator model, which it concluded to be rife with "unarticulated and questionable assumptions."²⁴

Compliance indicators

Of great relevance to the present study are the nascent attempts to develop trafficking-related indicators that seek to measure or otherwise assess national compliance with a particular legal or policy regime. For example, a 2007 European Union manual provides a range of indicators to measure the outcomes of EU Member States' responses to trafficking, based on a set of "guiding principles" drawn from the international and regional anti-trafficking legal and policy framework.²⁵ In 2011, the Association of South East Asian Nations (ASEAN) went much further: issuing a *Progress Report on Criminal Justice Responses to Trafficking in Persons* that set out detailed compliance indicators against "seven essential components of an effective criminal justice response to trafficking."²⁶ In his introduction to the Report, the ASEAN Secretary-General confirmed that the indicators are based on: "a combination of international, regional and national law and policy as well as accepted good practices within and beyond the ASEAN region."²⁷

While efforts to develop composite indicator sets that reference a range of standards are likely to continue, it is reasonable to expect that future attempts to shape compliance indicators in the area of trafficking will be anchored to one or both of the two major treaties: the European Trafficking Convention and the UN Trafficking Protocol. Accordingly, the structure and functioning of their respective monitoring mechanisms and the steps that have been taken thus far towards development of compliance indicators, deserve brief consideration.

The European Trafficking Convention

The European Trafficking Convention establishes a compliance system comprising two bodies: a technically oriented Group of Experts against trafficking in human beings (GRETA), and a more politically oriented Committee of the Parties, which is linked directly to the Council of Europe's Committee of Ministers. Each state

²⁴ NORAD, "Evaluation of the International Organization for Migration and its Efforts to Combat Trafficking" (2010), 56.

²⁵ European Commission, *Measuring Responses to Trafficking in the European Union: An Assessment Manual* (European Commission, 2007).

²⁶ Association of South East Asian Nations (ASEAN), "Progress Report on Criminal Justice Responses to Trafficking in Persons in the ASEAN Region" (hereinafter ASEAN, "Progress Report") (2011), iii. Note that one of the authors (Gallagher) was involved in the development of this Report.

²⁷ ASEAN, "Progress Report" (n. 26), iii.

party is subject to evaluation every four years, required to answer a questionnaire developed by GRETA to ascertain compliance with certain aspects of the Convention.²⁸ State responses can be supplemented by information from other sources, including visits and hearings. Draft reports, that include recommendations for action, are provided to the subject states for review and comment before being finalized and made public.²⁹ The Committee of Parties can step in at this point to request that a state party take certain measures to implement GRETA's conclusions.

It is through the questionnaire that GRETA has developed its indicators of compliance. These indicators are extensive and detailed, a reflection of the comprehensive nature of the Convention itself. However, while based on, and referring to, specific provisions of the Convention, they go much further. For example, indicators of compliance with respect to a state party's obligation to *prevent* trafficking extend beyond the strict requirements of the Convention to include consideration of the budget allocation for prevention efforts and the extent to which prevention strategies and campaigns are subject to impact assessment.³⁰

The inaugural evaluation round commenced in February 2010, with an extensive questionnaire eliciting information on the extent to which national laws, institutions, and processes conform to the relatively detailed standards prescribed by the Convention. By early 2012, country reports for Albania, Austria, Bulgaria, Croatia, Cyprus, Denmark, and the Slovak Republic had been made public. The evaluation process is quickly gaining momentum and it should be possible, within the next year or so, to make a preliminary assessment of the character and impact of this particular compliance regime.

The UN Trafficking Protocol

Unlike the European Trafficking Convention, the UN Trafficking Protocol does not establish a formal body to oversee its implementation. Article 32(1) of its parent instrument, the Organized Crime Convention, establishes a Conference of the Parties (COP) "to improve the capacity of States to combat transnational organized

²⁸ Council of Europe, Group of Experts on Action against Trafficking in Human Beings (GRETA), Rules of Procedure for Evaluating Implementation of the Council of Europe Convention on Action against Trafficking in Human Beings by the Parties, THB-GRETA 2009/3, June 17, 2009 (hereinafter GRETA, Evaluation Rules of Procedure).

²⁹ European Trafficking Convention (n. 3), art. 38. See also European Trafficking Convention Explanatory Report, paras 363–9; Council of Europe, Directorate General of Human Rights and Legal Affairs, Workflow of the Monitoring Mechanism of the Council of Europe Convention on Action Against Trafficking in Human Beings, THB-INF(2009)3, available at <http://www.coe.int/t/dghl/monitoring/trafficking/Source/Workflow_en.pdf> (accessed February 2, 2010); and GRETA, Evaluation Rules of Procedure (n. 28).

³⁰ Council of Europe, Group of Experts on Action against Trafficking in Human Beings (GRETA), Questionnaire for the evaluation of the implementation of the Council of Europe Convention on Action against Trafficking in Human Beings by the parties, First evaluation round, GRETA(2010)1 rev, February 1, 2010.

crime and to promote and review the implementation of [the] Convention.”³¹ The COP is empowered to request and receive information on states parties’ implementation of the Convention and its protocols and to make recommendations for improvement.³² The information received is compiled by the Secretariat into narrative reports that seek to provide an overview of trends in national responses.³³ In relation to the Trafficking Protocol, the COP has focused on collecting information regarding states parties’ efforts to fulfill specific obligations, for example, criminalizing trafficking and repatriating victims.

The COP reporting procedure has proven to be a crude mechanism for promoting or measuring compliance.³⁴ Few states parties provide full and complete information and many do not bother to respond at all. A Working Group established to improve COP procedures with respect to the Protocol³⁵ has made useful progress in identifying implementation priorities and challenges and even elaborating on the substantive content of some key norms. But it has not gone so far as to establish—or even call for—specific indicators of compliance. Recent efforts to strengthen the compliance regime around the Convention and its protocols have met with resistance. States have made clear that while they are willing to consider establishing a mechanism that is “transparent and efficient,” there are limits on what would be acceptable; for example, the focus of any such mechanism should not be on compliance per se, but rather on helping to develop national policies for implementation, as well as technical assistance and international cooperation initiatives.³⁶ In what appears to be a direct reaction to the unilateral monitoring and sanctioning regime established by the United States (discussed

³¹ To that end, the Convention envisages that the Conference of Parties (COP) will facilitate training, technical assistance, and related activities; facilitate information exchange between states parties on patterns and trends in organized crime and successful responses; and cooperate with relevant intergovernmental and non-governmental organizations. GRETA, Evaluation Rules of Procedure (n. 28), arts 32(3)(a)–32(3)(c).

³² Conference of Parties to the United Nations Convention on Transnational Organized Crime, Decision 1/5, Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, supplementing the United Nations Convention against Transnational Organized Crime, reproduced in Conference of Parties to the United Nations Convention on Transnational Organized Crime, Report of the Conference of Parties to the United Nations Convention on Transnational Organized Crime on its first session, held in Vienna from June 28 to July 8, 2004, September 23, 2004, UN Doc. CTOC/COP/2004/6.

³³ For the most recent report, see Conference of the Parties to the United Nations Convention against Transnational Organized Crime, Implementation of the Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children, Supplementing the United Nations Convention against Transnational Organized Crime: Consolidated Information Received From States for the Second Reporting Cycle, September 9, 2008, UN Doc. CTOC/COP/2006/6/Rev.1.

³⁴ Anne T. Gallagher, “Improving the Effectiveness of the International Law of Human Trafficking: A Vision for the Future of the U.S. Trafficking in Persons Reports,” *Human Rights Review* 12 (2011): 381 (hereinafter Gallagher, “TIP Reports Review”).

³⁵ Conference of Parties to the United Nations Convention on Transnational Organized Crime, Decision 4/4, Trafficking in Human Beings, reproduced in Conference of Parties to the United Nations Convention on Transnational Organized Crime, Report of the Conference of Parties to the United Nations Convention on Transnational Organized Crime on its fourth session, held in Vienna from October 8 to 17, 2008, December 1, 2008, UN Doc. CTOC/COP/2008/19, 11.

³⁶ Conference of the Parties to the United Nations Convention against Transnational Organized Crime, Report on the Meeting of Experts on Possible Mechanisms to Review Implementation of the

below),³⁷ states parties have declared—and major anti-trafficking organizations have echoed³⁸—that any oversight mechanism established under the Convention (and thereby the Protocol) must also be: “non-intrusive, impartial, non-adversarial, non-punitive and flexible. In addition, it should not criticize or rank States or regions but rather contribute to problem-solving and respect the sovereignty of States.”³⁹ It can be expected that any official efforts aimed at improving the effectiveness of the assessment process, for example through the formulation of compliance indicators, will meet strong and widespread resistance. The technical difficulty of constructing and applying appropriate measurement tools should also not be underestimated, as potentially illustrated by recent efforts of academic economists to develop an indicator-based compliance assessment model for the Protocol.⁴⁰

Outside the COP, there have been several attempts to harness the political and legal weight of the Trafficking Protocol through development of assessment indicators. The American Bar Association, for example, has produced a Human Trafficking Assessment Tool, which seeks to measure states parties’ de facto and de jure compliance with the Protocol.⁴¹ In 2009, the UNODC released a *Global Report on Human Trafficking* (UNODC Report), which collected and analyzed a range of data from 155 states regarding a handful of “indicators” in three subject areas: legislative and administrative frameworks; criminal justice; and victim services. Clearly anticipating political sensitivities, the Report was careful not to characterize these as compliance indicators relative to the Trafficking Protocol; rather, they were presented as “particularly relevant for cross-national data collection

United Nations Convention against Transnational Organized Crime held in Vienna on September 30, 2009, October 14, 2009, UN Doc. CTOC/COP/WG.1/2009/3, para. 21.

³⁷ See further the discussion below in the second part of this chapter.

³⁸ See, e.g., Global Alliance against Trafficking in Women and Others, “Statement on a Monitoring Mechanism for the United Nations Convention against Transnational Organized Crime and Each of the Protocols Thereto with Specific Attention to the Protocol to Prevent, Suppress and Punish Trafficking in Persons (the Human Trafficking Protocol),” October 13, 2009, available at <http://www.gaatw.org/statements/Statement_on_a_Monitoring_Mechanism-COPS08.pdf> (accessed February 2, 2010) (“[a review mechanism should be] non-punitive, non-adversarial, non-conditional and non-ranking”).

³⁹ Conference of the Parties to the United Nations Convention against Transnational Organized Crime, Report on the Meeting of Experts on Possible Mechanisms to Review Implementation of the United Nations Convention against Transnational Organized Crime held in Vienna on September 30, 2009, October 14, 2009, UN Doc. CTOC/COP/WG.1/2009/3, para. 22.

⁴⁰ In a recent series of papers, researchers from the Department of Economics, University of Göttingen, develop and apply a highly complex and obscure index of state party compliance with the “prevention, protection and prosecution” obligations of the Protocol (a typology not found in the Protocol itself but based on the structure of the US TIP Reports). The authors are not fazed by the difficulty in securing reliable information to feed into the index: their only sources of information on state responses to trafficking, adopted without comment or qualification, are the TIP Reports and 2009 UNODC Report. Evidence from a New Index,” Mimeo, Georg Seo-Young Cho, Axel Dreher, and Eric Neumayer, “The Spread of Anti-trafficking Policies,” Mimeo, Georg—August University of Göttingen, Germany (2011), and Seo-Young Cho and Krishna Chaitanya Vadlamannati, “Compliance for Big Brothers An Empirical Analysis on the Impact of the Anti-trafficking Protocol,” Mimeo, Georg-August University of Göttingen, Germany (2011).

⁴¹ American Bar Association, Rule of Law Initiative, Human Trafficking Assessment Tool (2005), available at <http://www.abanet.org/rol/publications/human_trafficking_assessment_tool.shtml>.

and exchange.”⁴² Unfortunately, the information presented in the report was of a low quality: unverified, uneven, and seriously incomplete.⁴³ UNODC did not seem particularly concerned by these weaknesses, confidently asserting that “[a] poor indicator is better than no indicator as long as it is not represented as more than it is.”⁴⁴ Readers who may have been unsettled by this pronouncement were reassured that: “over time, the collection of information from so many different perspectives can, in aggregate, make up for many of the deficiencies in the data itself.”⁴⁵

UNODC’s only other foray into the world of trafficking-related indicators came in 2010 with the release of the *International Framework for Action to Implement the Trafficking in Persons Protocol*. The Framework for Action was promoted as a guide to states parties in implementing their obligations under the Protocol.⁴⁶ However, while the Trafficking Protocol is clearly the inspiration and major reference point, the document draws on a wide range of interpretative texts, policy documents, guidelines, and other materials in formulating overall guiding principles, as well as indicators and implementation measures.

Case study: US TIP indicators

The scope, influence, and authority of the various tools and mechanisms outlined above pale in comparison beside the unilateral, indicator-based compliance mechanism created by the US Congress to oversee the response of other countries to trafficking. As described in greater detail below, each year, the US State Department publishes a Trafficking in Persons Report (TIP Report), in which it applies a set of US-defined “minimum standards” to assess and rank the efforts of every country to eliminate trafficking, including, from 2009, its own. Those countries falling into the bottom tier are potentially subject to a range of economic sanctions. Despite careful avoidance of the language of indicators (except, as noted above, with reference to *indications* of trafficking), the evaluative criteria and rankings system at the heart of the US compliance regime identifies it as a classic example of the production and use of indicators as means of exercising authority and power. Assessments are based on a basket of largely qualitative measures (“minimum standards” and “criteria”) that are then used to construct a single and all-important numerical ranking. This ranking facilitates comparisons between countries as well as the tracking of changes in individual performance.

This compliance mechanism has enabled the US government to assume a central role in global anti-trafficking efforts, effectively displacing the international and regional laws and machinery outlined above. Most critically in the present context,

⁴² UNODC Global Report (n. 4), 72–3.

⁴³ Gallagher, “TIP Reports Review” (n. 34).

⁴⁴ UNODC Global Report (n. 4), 13.

⁴⁵ Ibid.

⁴⁶ United Nations Office on Drugs and Crime (UNODC), *International Framework for Action to Implement the Trafficking in Persons Protocol* (2009).

the indicators at the heart of the Reports have played a crucial role in both knowledge production and governance: shaping perceptions about what trafficking is; its causes and its consequences; how governments should respond; and what a positive or inadequate response looks like. The results of this effort have been predictably mixed. On the positive side, there can be little doubt that the US TIP compliance mechanism has encouraged greater state action against trafficking than would otherwise have been the case. As one of the present authors recently noted, “[w]ithout the Reports, our collective knowledge of trafficking-related exploitation would likely be less; individual governments would likely have greater control over the flow of information that properly belongs in the public domain; and even the most egregious failure on the part of a state to deal with trafficking-related exploitation would likely come at little reputational or other cost.”⁴⁷ Those important achievements need to be balanced against some serious risks: the risk that this unilateral compliance regime will damage international law and international institutions; and the risk of unanticipated negative consequences including unfair trials, inappropriate sentencing, and discriminatory action against vulnerable groups such as migrant workers.⁴⁸

Overview of TIP Reports and sanctions mechanism

The US Victims of Trafficking and Violence Protection Act of 2000 (TVPA) was signed into law on October 11, 2000, two months before the adoption of the UN Trafficking Protocol. During the drafting process, the question of whether the TVPA should include an assessment and sanctions regime for other countries was hotly debated.⁴⁹ The Clinton Administration was firmly opposed to the development of a unilateral monitoring mechanism, its officials arguing that making “an international pariah” out of a deficient country would almost certainly undermine the UN Trafficking Protocol’s goal of fostering international cooperation. It was further argued that the sanctions threat could cause governments to downplay their trafficking problems—thus undercutting critical transnational data collection on trafficking patterns and practices—and hinder intra-national cooperation by causing governments and local populations to view as a threat the work of NGOs to track and expose the problem within the country.⁵⁰

Notwithstanding these concerns, the TVPA’s congressional sponsors believed that the success of efforts to prevent trafficking into the United States depended heavily on the actions of other countries. Accordingly, the TVPA establishes a

⁴⁷ Gallagher, “TIP Reports Review” (n. 34).

⁴⁸ Ibid. Janie Chuang, “The United States as Global Sheriff: Using Unilateral Sanctions to Combat Human Trafficking,” *Michigan Journal of International Law* 27 (2006): 437 (hereinafter Chuang, “Global Sheriff”).

⁴⁹ International People Trafficking: Hearing Before the S. Foreign Relations Comm. Subcommittee on Near Eastern and South Asian Affairs, 106th Cong. (2000) (Statement of Bill Yeomans, Chief of Staff, Department of Justice Civil Rights Division) (hereinafter Yeomans testimony). See also Janie Chuang, “Global Sheriff” (n. 48).

⁵⁰ Yeomans testimony (n. 49).

system whereby the State Department issues annual TIP Reports describing “the nature and extent of severe forms of trafficking in persons” and assessing governmental efforts across the world to combat such trafficking against criteria established by US law.⁵¹ The indicators upon which these assessments are made and the process for translating this information into rankings and potential sanctions are detailed below.

The indicators

The TVPA measures a country’s efforts to eliminate trafficking against a set of four meta-indicators (“minimum standards”):

- (1) The government should prohibit and punish acts of severe forms of trafficking in persons.
- (2) For sex trafficking involving force, fraud, coercion, or in which the victim is a child, or of trafficking which involves rape, kidnapping, or death, the government should prescribe punishment commensurate with that for grave crimes.
- (3) For the knowing commission of any act of a severe form of trafficking, the government should prescribe punishment that is stringent enough to deter and that reflects the heinous nature of the offense.
- (4) The governments should make serious and sustained efforts to eliminate severe forms of trafficking in persons.⁵²

In evaluating “serious and sustained efforts to eliminate severe forms of trafficking,” the TVPA provides a comprehensive set of (predominantly qualitative) indicators, summarized as follows: (i) whether the government vigorously investigates, prosecutes, and punishes trafficking; (ii) whether it protects victims and encourages their participation in the investigation and prosecution process; (iii) whether it has adopted preventative measures such as public education, birth registration, control of nationals deployed abroad in peacekeeping and similar operations, and measures aimed at preventing forced labor and child labor; (iv) whether it cooperates with other governments in investigations and prosecutions; (v) whether it extradites (or is attempting to enable extradition of) traffickers; (vi) whether it monitors migration patterns for evidence of trafficking and responds to such evidence in an appropriate manner; (vii) whether it investigates, prosecutes, and takes appropriate measures against the involvement of public officials in trafficking; (viii) whether the percentage of victims of trafficking that are non-citizens is insignificant; (ix) whether the government has taken measures to address demand for trafficking related to commercial sex acts and involvement of nationals in sex tourism; (x) whether it has

⁵¹ The TIP Reports did not emerge in a legal or policy vacuum but form part of an established tradition of US congressional oversight of the actions of other countries with respect to issues considered to be of particular political significance. Gallagher, “TIP Reports Review” (n. 4).

⁵² TVPA (n. 6), § 108(a).

achieved appreciable progress as measured against the previous year's assessment; and (xi) whether it has monitored and provided information to the US government on its national response to trafficking.⁵³

The minimum standards and their associated indicators have been revised and expanded with each Reauthorization of the TVPA (i.e., in 2003, 2005, and 2008). Some revisions sought to bring within the scope of the Report particular forms of trafficking that have traditionally been under-reported and under-addressed—for example, involvement of government contractors and peacekeeping personnel in trafficking activities.⁵⁴ Other revisions reflect an effort to ensure that the actions taken by a government are meaningful with respect to the goal of eradicating trafficking. For example, the prosecution criteria expanded from initially requiring government data regarding investigations and prosecutions to also including consideration of convictions and sentences,⁵⁵ and consideration of whether severely reduced or suspended sentences indicate serious efforts to combat trafficking.⁵⁶ Still other revisions to the US minimum standards were directed to making the assessment process more effective by addressing barriers to meaningful scrutiny. For example, recognizing many governments' failure to systematically collect data on trafficking patterns and government responses, the 2003 TVPRA turned the availability of data into an indicator in itself, and further added the criteria that governments conduct and make publicly available periodic assessments of their own progress in combating trafficking.⁵⁷

GTIP asserts that individual country performance against the established indicators is based on qualitative and quantitative data compiled from a wide variety of sources, including US embassies, government officials, NGOs and international organizations, published reports, research trips to every region, and information submitted directly to its Office.⁵⁸ The main vehicle for information collection is a questionnaire that closely tracks the indicators attached to the TVPA "minimum standards."⁵⁹ The Washington-based GTIP office sends the questionnaire to a focal point in each US embassy, and it is the embassy's responsibility to secure information from the appropriate sources. Many questions, particularly those seeking hard data (e.g., numbers of prosecutions) are directed only at the government under assessment. Others (e.g., quality of government-provided victim services) are directed to service providers and those who may have information or insight on such matters. Embassy officials may also consult with international organizations, journalists, academics, and trafficking survivors to gather the requisite data.⁶⁰ GTIP sends an abbreviated questionnaire to NGOs to solicit information regarding the

⁵³ TVPA (n. 6), § 108; 2003 TVPRA (n. 6), § 6(d); 2005 TVPRA (n. 6), § 104(b); 2008 TVPRA (n. 6), § 106.

⁵⁴ 2005 TVPRA (n. 6), § 104(b)(1)(A), (B).

⁵⁵ 2003 TVPRA (n. 6), § 6 (d)(1)(A).

⁵⁶ 2008 TVPRA (n. 6), § 106(2)(1).

⁵⁷ 2003 TVPRA (n. 6), § 6 (d)(1)(B), (d)(2)(B).

⁵⁸ See, e.g., 2011 TIP Report (n. 9), 11 (outlining methodology for data collection).

⁵⁹ See, e.g., Department of State, Request for Information for the 2011 Trafficking in Persons Report, in *Federal Register*, Vol. 76, No. 7 (January 11, 2011), 1660–3.

⁶⁰ 2010 TIP Report (n. 4), 20.

government's efforts to meet the US minimum standards. A copy of that questionnaire is not publicly available, nor is information on how NGOs are selected to participate in this process. GTIP actively encourages NGOs and other interested parties to submit information via email to <tipreport@state.gov>.⁶¹ It is unclear how and to what extent these supplementary sources of information are evaluated and used.

According to GTIP officials, the response to the core questionnaire, prepared by embassy staff, is used as a basis for discussion between the embassy and GTIP, focusing particularly on substance, implications, and reliability of the information collected. The embassy may be requested to secure additional data. If information from the government seems doubtful (e.g., an unusually large number of prosecutions), GTIP will seek to secure corroboration from other sources (e.g., a national NGO). Importantly, there is no established data verification procedure: GTIP retains full discretion to determine the need for and level of verification of any information received or used. When information from different sources conflicts (e.g., government vs NGO regarding the level of trafficking-related government corruption), embassy staff may meet with relevant parties to determine which, in their view, is the more accurate version.⁶² In determining how much weight to give to NGO findings, GTIP considers whether the NGO has known competency with respect to the issue at hand. It appears that a similar weighting system is used when assessing the reliability of information coming from the government under assessment.

Tier rankings and sanctions mechanism

Using categories established by the TVPA, the TIP Report ranks countries according to their level of compliance with the US minimum standards. Tier One is for countries in full compliance with the minimum standards set out above, Tier Two for countries making an effort but not yet fully compliant, and Tier Three for those countries that are failing on both counts.⁶³ Subsequent amendments laid the ground for creation of an additional category, "Tier Two Watch List," applied to countries that, owing to the severity of the problem or failure to provide evidence of progress, are considered to be on the lower edge of Tier Two classification.⁶⁴ Tier Two Watch List countries are subject to special scrutiny and, in the absence of a special presidential waiver, are downgraded to Tier Three after two consecutive years on the Watch List.⁶⁵ Under the TVPA and its various amendments, the President is

⁶¹ See, e.g., 2010 TIP Report (n. 4), 19.

⁶² Clearly, this additional step would only be available in those countries with a robust civil society that is active in the area of counter-trafficking and not subject to government control or intimidation.

⁶³ TVPA (n. 6), § 110.

⁶⁴ GTIP unilaterally added the Tier Two Watch List to the rankings in response to the 2003 TVPRA's introduction of a "special watch list" of countries that would be subject to special scrutiny each year. This occurred despite the clear lack of congressional intent to add another tier to the rankings. 2003 TVPRA (n. 6), § 6(e).

⁶⁵ 2008 TVPRA (n. 6), § 107(a). The criteria for the presidential waiver in effect function as additional indicators for the assessment of whether these Tier Two Watch List countries should be downgraded and thereby subject to the threat of sanctions. The waiver is justified if: (1) the country has

authorized to deny the provision of non-humanitarian, non-trade-related assistance to any Tier Three country, i.e., any government that does not comply with the minimum standards and that is not making significant efforts to bring itself into compliance. In addition, such countries risk US opposition to their seeking and obtaining funds from multilateral financial institutions, including the World Bank and the IMF.

The annual Reports are used as a basis for determining whether, and to what extent, sanctions are to be imposed or assistance provided. Countries have 90 days after issuance of the TIP Report to work with the State Department to remedy identified failings and avoid sanctions. The President has broad discretion to waive sanctions—i.e., where doing so would be in the “national interest,” promote the purposes of the TVPA, or avoid significant adverse effects on vulnerable populations.⁶⁶ This waiver power has been regularly exercised and penalties have generally only been imposed on those countries already under US economic sanctions or with which the United States has little economically and strategically at stake.⁶⁷ The 2012 decision by President Obama to waive sanctions against Burma in recognition of recent, important domestic reforms appears to confirm that the granting or withholding of a waiver is a discretionary political act that owes little or nothing to the situation of trafficking in the target country.

The centerpiece of each TIP Report is a table that sets out the tier ranking of all countries assessed. The justification for tier rankings is provided in individual “country narratives,” which, over the course of several pages, assess each government’s efforts to comply with the US minimum standards. The narrative for each country assessment is organized into three separate sections (prosecution, protection, and prevention) that closely track the indicators summarized above. Each narrative concludes with a list of country-specific recommendations for improving compliance with the US minimum standards, and includes a graph charting the country’s tier rankings since its first inclusion in the Report. Anecdotal evidence gained through the authors’ work in the field and an examination of official responses to the Reports suggests that it is the tier rankings themselves that are considered critical. Governments appear to care far more about their actual grade than they do about the substantive analysis, findings, and recommendations contained in the relevant country assessment.

a written plan to begin making significant efforts to bring itself into compliance with the minimum standards for the elimination of trafficking; (2) the plan, if implemented, would constitute making such significant efforts; and (3) the country is devoting sufficient resources to implement the plan. Ibid.

⁶⁶ TVPA (n. 6), § 110(d).

⁶⁷ Countries sanctioned to date include Burma, Zimbabwe, North Korea, Syria, Equatorial Guinea, Venezuela, and Cambodia. Cuba, Eritrea, Iran, Syria, Fiji, Sudan, Liberia, and Madagascar, which had not received financial assistance prior to the sanctions determination, were denied funding for participation in educational and cultural exchanges. See Presidential Determination with Respect to Foreign Governments’ Efforts Regarding Trafficking in Persons (2003–2010). Anecdotal information suggests that the greater impact of the sanctions mechanism may be felt in the multilateral lending context, with the promise of US withholding of support apparently deterring bank officials from even *proposing* aid projects in the sanctioned countries.

Nature and quality of knowledge production through the US TIP indicators

While the US indicator-based compliance mechanism appears relatively straightforward, closer examination of how the indicators are applied in practice reveals practices of measurement that cast doubt on the objectivity of the reporting process and thereby the quality of the information generated. For example, GTIP exercises broad discretion as to how and which indicators are to be applied. Moreover, the process by which the TIP Report is produced reveals structural barriers that can undermine the reliability of the data presented and invariably subject the TIP Report contents to a political litmus test.

Discretionary application of indicators

GTIP enjoys broad discretion with respect to how it applies the indicators embodied in the minimum standards. The four “meta-indicators,” for example, are not considered equally. The first three of these relate to prohibition and penalization of trafficking. It is only under the final meta-indicator (“the government should make serious and sustained efforts to eliminate severe forms of trafficking”) that the full range of victim protection and support matters fall. Even more so than the other three, this particular indicator is capable of broad interpretation and, historically, has been used to promote a wide range of actions and responses. It is also relevant that the detailed criteria attached to the fourth indicator are framed as indicia that “should be considered.” There is no language suggesting that the list of relevant concerns is exhaustive or that each criterion should be weighted equally. Criteria that have fallen out of favor (or perhaps never made sense to begin with) are passed over without comment.⁶⁸

Moreover, the question of what is being measured cannot be fully answered with reference to the formal indicators of compliance summarized above. Rather, a closer examination of the TIP Reports reveals the application of additional and different criteria that go well beyond the legislatively mandated “minimum standards.” These “shadow” indicators do not yet figure largely in the formal country assessments, which are, as noted previously, carefully scripted according to predetermined, legislatively approved categories. Shadow indicators are, at present, most evidently at play in the expanding TIP Report introduction, which, in addition to highlighting various forms and examples of trafficking-related exploitation, is also used to present the US government’s understanding of the causes and consequences of trafficking, and to provide examples of what the State Department considers to be appropriate government responses to the problem. The 2011 Report, for example, explores issues as diverse as forced child labor and child

⁶⁸ For example, a close review of the individual country narratives over the past few years reassuringly confirms that an additional, inexplicable criterion introduced by the 2005 TVPRA of “[w]hether the percentage of victims of severe forms of trafficking in the country that are non-citizens of such countries is insignificant,” has generally been ignored. 2005 TVPRA (n. 6), § 6(d)(3).

soldiers; trafficking by government contractors; exploitative labor recruitment practices; and the need for victim-centered interviewing techniques.⁶⁹ It is in the context of this general presentation of issues that the US government encourages all countries to “make serious and sustained efforts that rise above the minimum standards and the . . . [Trafficking] Protocol [standards] to put forth a truly comprehensive governmental response.”⁷⁰

Examples of shadow indicators from the 2011 report include whether the government has: engaged in proactive victim identification measures; funded and partnered with NGOs; ensured that worker rights are protected and supply chains are free of forced labor; and provided victims with options for immigration relief and basic services including shelter, food, medical care, psychological counseling, legal aid, and work authorization.⁷¹ The recommendations that relate to these indicators are all broader and far more specific than the victim protection provisions of the US minimum standards—or indeed, of international legal obligations. While their overt expression is typically confined to the general introductory sections of the report, the 2011 TIP Report explicitly notes their influence on tier rankings and narratives.⁷² The issue of shelter for victims of trafficking, for example, is not referred to in the minimum standards. Increasingly however, the TIP Report assessments and recommendations address shelter availability; the scope and quality of shelter services; the voluntary or coerced nature of shelter provision; and the willingness of governments to fund shelters.⁷³

The politics of measurement

A close examination of the process by which the TIP Reports are produced reveals important structural barriers to objectivity and reliability of data collection, as well as political filtering of the information ultimately included in the final reports.

The issue of data objectivity and reliability is particularly troublesome. Despite protestations to the contrary and a marked improvement in data quality over the past several years, much of the information contained in the TIP Report continues to be second-hand, unverified, and often unverifiable. This is partly the result of broader, systemic obstacles related to the collection and validation of trafficking-related data.⁷⁴ However, it also reflects the Washington-centric nature of the reporting process, as well as limitations on GTIP resources. As of mid-2012, only 13 GTIP staff were responsible for assessing the approximately 180 countries covered in the annual TIP Report. Even with strong embassy support, this small group of officials does not have the capacity to conduct substantial primary research

⁶⁹ 2011 TIP Report (n. 9), 7–8, 19, 21, 26–9, 31–6.

⁷⁰ 2011 TIP Report (n. 9), 17.

⁷¹ 2011 TIP Report (n. 9), 11–13, 24–9, 40–2, 44–5.

⁷² 2011 TIP Report (n. 9), 11.

⁷³ 2011 TIP Report (n. 9), 63 (Afghanistan), 65 (Albania), 81 (Bahrain), 119 (Chile), 190 (India), 207 (Japan), 221 (Kuwait), 239 (Macedonia), 245 (Malaysia), 280 (Nigeria).

⁷⁴ See further, Gallagher and Surtees, “Measuring the Impact and Effectiveness of Anti-Trafficking Interventions in the Criminal Justice Sector” (n. 21).

or to corroborate the claims or findings of others. This situation differs from a similar congressional oversight mechanism of the State Department—the annual *Country Reports on Human Rights Practice*—for which country assessments are prepared in-country, with the relevant Washington-based bureau of the State Department exercising only a light editing role. While vesting the assessment process in a core group of individuals at GTIP builds expertise and ensures high levels of consistency, it also limits their ability to fully contextualize the data collected for the country assessments. Deep debates over the very definition of trafficking and the range of practices that label encompasses—discussed below in the third part of this chapter—render context critical to proper identification of the phenomenon being measured, not to mention assessment of the “anti-trafficking” measures taken in response.

Vetting of draft narratives and proposed rankings by other (non-GTIP) offices subjects the document to political litmus tests that invariably compromise the objectivity and thoroughness of the country assessments. It is perhaps unrealistic to expect that embassies and Washington-based regional bureaus at the State Department ignore broader political ramifications of TIP Report assessments for US foreign relations. In practice, disputes over the language in the draft report and/or provisional tier ranking are subject to negotiation—which, as has been noted by the Government Accountability Office, may involve some “horse-trading,” whereby GTIP “agrees to raise some countries’ tier rankings in exchange for lowering others.”⁷⁵ If negotiations fail to reach consensus, the matter is sent further up the political chain to the Under-Secretary of Global Affairs for resolution. It is only once cleared by the relevant political offices of the State Department, that the final country assessments and tier rankings are packaged and presented within the annual TIP Report.⁷⁶

The malleability of this indicator-based, yet highly politicized system of assessment and measurement is usefully illustrated with reference to the TIP Report’s treatment of the issue of domestic worker trafficking by diplomats.⁷⁷ There can be little doubt that this manifestation of trafficking is particularly problematic, not least because of the immunity from prosecution and civil suit that most diplomats enjoy under international law.⁷⁸ The political sensitivities of naming and shaming government officials undermine GTIP efforts to implement the TVPA requirement that the TIP Report address “[w]hether [a] government . . . vigorously investigates and prosecutes public officials who participate in or facilitate [trafficking].”⁷⁹ Indeed, in at least one case, it appears that political factors explain the otherwise unfathomable omission of a known situation of diplomat-related trafficking from

⁷⁵ United States Government Accountability Office, GAO Human Trafficking (n. 19), 33. As the Report notes, disagreements between GTIP and the regional bureaus may be expected. GTIP is focused solely on anti-trafficking issues, whereas the Regional Bureaus are required to consider and address a wide range of issues affecting bilateral relations between the United States and another government. *Ibid.*, 32–3.

⁷⁶ Interviews with staff of the US State Department Office to Monitor and Combat Trafficking in Persons (September 2, 2010).

⁷⁷ Janie Chuang, “Achieving Accountability for Migrant Domestic Worker Abuse,” *North Carolina Law Review* 88 (2010): 1627.

⁷⁸ Vienna Convention on Diplomatic Relations, April 18, 1961, 500 UNTS 95.

⁷⁹ TVPA (n. 6) § 108(b) (codified as amended at 22 USC § 7106(b)(7)(2006)).

the relevant country narrative.⁸⁰ Despite a congressional rebuke and explicit direction to include such cases in the TIP Report,⁸¹ the 2010 and 2011 TIP Reports continued to omit references even to *adjudicated* diplomatic trafficking cases.⁸²

Implications for global governance

The US TIP compliance mechanism has enabled the United States to assume the leading role in global anti-trafficking governance. Through the Reports, the United States has exercised a disproportionate influence over normative development in international anti-trafficking law and policy, as well as over the direction of national responses. The ways in which this has occurred, and the implications for the future, are discussed further below.

1. Legal and normative implications

i. Correlation with international norms

From the perspectives of international law and policy, the very existence of the US TIP compliance mechanism is cause for deep concern: the performance of states with respect to trafficking is currently being assessed, not with reference to the international rules that they have together helped to develop and freely accepted, but against criteria created and imposed by US bureaucrats and politicians.⁸³

While there is substantial overlap between the US criteria and relevant international law, the two regimes are not identical, and the TVPA indicators do not precisely correspond to the rights and obligations that arise through international law. This anomaly is clearly felt by the authors of the Report, whose treatment of international law has been inconsistent and ambiguous. While references to international norms and standards have increased in recent reports, such references are generally superficial. In its introductory section, the Report has occasionally referred to international instruments, noting, for example, that the Trafficking Protocol definition extends to trafficking for labor exploitation as

⁸⁰ In a 2008 case, a migrant domestic worker won a \$1 million judgment against her US-based Tanzanian diplomat-employer for subjecting her to involuntary servitude and forced labor (*Mazengo v Mzengi*, 542 F. Supp.2d 96, 97–8 (DDC 2008)). The worker's lawyers have been lobbying the State Department to mention the case in the TIP Reports in order to assist her to collect on the judgment. To date, no mention of the case has yet been made, despite publicity (Sarah Fitzpatrick, "Diplomatic Immunity Leaves Abused Workers in Shadows," *Washington Post*, September 20, 2009, A4). According to news reports, the defendant-employer in the *Mazengo* case was appointed to be a direct adviser to the President of Tanzania. Benjamin Skinner, "Modern-Day Slavery on D.C.'s Embassy Row?" *Time Magazine*, June 14, 2010, available at <<http://www.time.com/time/nation/article/0,8599,1996402,00.html>> (last visited September 8, 2010).

⁸¹ 2008 TVPRA (n. 6), §§ 202–3. Consolidated Appropriations Act of 2010, H.R. 3288, 111th Cong. § 7034(s) (2009).

⁸² 2010 TIP Report (n. 4), 318–20; 2011 TIP Report (n. 9), 349–51.

⁸³ Gallagher, "TIP Reports Review" (n. 34).

well as for sexual exploitation,⁸⁴ and referencing the Trafficking Protocol and the European Trafficking Convention in support of general obligations, including criminalization of trafficking and protection of victims.⁸⁵ The Report includes a table of ratifications covering a small group of relevant treaties, including the ILO conventions on forced and child labor.⁸⁶ However, the text of the TIP Report does not refer to the substantive standards contained within these treaties in its general commentary or analysis of country practices. Individual country assessments typically only consider whether that country is party to the Trafficking Protocol, and if not, recommending accession.⁸⁷ There is no attempt to evaluate the quality and legality of state responses to trafficking against, for example, international legal standards governing the right to a fair trial and the obligation to provide access to remedies.

In an apparent response to criticism of GTIP marginalization of international law, the 2011 Report sets out a detailed explanation of the link between the TVPA and the Trafficking Protocol, stating that “the standards in the TVPA . . . are largely consistent with the framework for addressing trafficking set forth in the . . . Protocol, both in form and content.”⁸⁸ The analysis concludes that: “although each TIP Report presents assessments under American law, the standards they build from are firmly rooted in international law.”⁸⁹ This is an overdue and welcome affirmation of the importance of the international legal framework. However, the conclusion is not entirely convincing, being called into question by the apparent willingness of GTIP to manipulate key concepts to suit its own narrative. The definition of trafficking provides a useful illustration. While there are some important differences between the TVPA definition and that set out in the UN Trafficking Protocol, each contains essentially the same complex and, in some respects, ambiguous elements.⁹⁰ At the international level, the precise contours of the agreed definition remain unsettled.⁹¹ GTIP has begun to weigh into this process by advancing an

⁸⁴ See, e.g., US Department of State, Trafficking in Persons Report 2003, 135 (hereinafter 2003 TIP Report); US Department of State, Trafficking in Persons Report 2005, 9 (hereinafter 2005 TIP Report).

⁸⁵ See, e.g., US Department of State, Trafficking in Persons Report 2008, 27, 38 (hereinafter 2008 TIP Report); US Department of State, Trafficking in Persons Report 2009, 24 (hereinafter 2009 TIP Report); 2010 TIP Report (n. 4), 7; 2011 TIP Report (n. 9), 7, 16–17.

⁸⁶ 2010 TIP Report (n. 4), 360–4; 2011 TIP Report (n. 9), 398–402.

⁸⁷ See, e.g., 2011 TIP Report (n. 9), 133, 318, 321.

⁸⁸ 2011 TIP Report (n. 9), 16.

⁸⁹ *Ibid.*

⁹⁰ The US minimum standards target “severe forms of trafficking,” which the TVPA defines as:

a. sex trafficking in which a commercial sex act is induced by force, fraud, or coercion, or in which the person induced to perform such an act has not attained 18 years of age; or

b. the recruitment, harboring, transportation, provision, or obtaining of a person for labor or services, through the use of force, fraud, or coercion for the purpose of subjection to involuntary servitude, peonage, debt bondage, or slavery.

TVPA (n. 6), § 103(8). The TVPA separately defines “sex trafficking” as “the recruitment, harboring, transportation, provision, or obtaining of a person for the purpose of a commercial sex act.” TVPA (n. 6), § 103(9). For an analysis of the distinctions between this definition and the one set out in the Trafficking Protocol, see Gallagher, “TIP Reports Review” (n. 34).

⁹¹ Anne T. Gallagher, *The International Law of Human Trafficking* (Cambridge: Cambridge University Press, 2010) (discussing the Protocol definition in Chapter 1).

interpretation of the definition that extends the concept of trafficking to include *all* forms of forced labor.⁹² Unsurprisingly, this interpretation has been roundly rejected by the traditional guardians of the forced labor mandate, including the International Labor Organization.⁹³ Whether or not the GTIP interpretation can be sustained as a legal matter remains to be seen. Of greater relevance to the present discussion is the potential impact of a GTIP-led reorienting of the definition on the nascent international anti-trafficking legal framework. There is, in short, a strong likelihood that the parameters of the accepted definition of trafficking will change—not because most or all states agree that this is required, but because one state that seeks this change is in a position to wield disproportionate influence over the evolution of the norm.

GTIP's conflation of trafficking with slavery is another example of the language and norms of international law being manipulated to suit a particular advocacy agenda.⁹⁴ There is, of course, a link between the two concepts. The international legal definition of trafficking recognizes slavery as a potential exploitative *purpose* for which individuals may be trafficked.⁹⁵ However, the identification of trafficking as a *manifestation* of the international legal concept "*slavery*" is something else entirely and rightly contested.⁹⁶ Inaccurate and unreflective attempts to expand the concept of slavery operate to diminish the legal worth of a powerful international legal prohibition that has been enshrined in the legal systems of many states. There is indeed a need to clarify the substantive content of the international legal prohibition on slavery in light of current challenges. However, that task rightly belongs to the international community and should not be unilaterally assumed.

A final, important example of normative manipulation relates to the difficult and controversial issue of prostitution. There is no international legal prohibition on prostitution. Apart from the issue of child prostitution and the (undefined) notion of *exploitation of prostitution*, international law reflects a clear and widespread preference for leaving the issue of prostitution in the hands of individual states.⁹⁷

⁹² 2010 TIP Report (n. 4), 8; 2011 TIP Report (n. 9), 7–8.

⁹³ The ILO distinguishes trafficking from forced labor by emphasizing the movement aspect of the trafficking definition: "not all victims of forced labour are in this situation as a result of trafficking. . . . people who are coerced to work in their place of origin have not been considered in the ILO's own estimates of forced labour as trafficking victims." International Labor Organization, "Fighting Human Trafficking: The Forced Labour Dimensions" (ILO, 2008). Certain segments of the US government, including the Department of Labor and the Department of State Bureau of Democracy, Rights, and Labor, adopted this distinction until very recently, when the US government adopted a unified position equating all forms of forced labor with trafficking.

⁹⁴ See, e.g., 2010 TIP Report (n. 4), 5–7, 13–16; 2011 TIP Report (n. 9), 7, 16, 18–22, 26, 32. See also, Luis CdeBaca, "Ongoing Fight Against Modern Slavery," speech at the 2010 National Conference of Child Abuse Professionals, sponsored by the National District Attorneys Association, August 23, 2010, available at <<http://www.state.gov/g/tip/rls/rm/2010/143495.htm>>; Luis CdeBaca, "Tenth Anniversary of Landmark Efforts to Combat Modern Slavery," speech, June 11, 2010, available at <<http://www.state.gov/g/tip/rls/rm/2010/142569.htm>>.

⁹⁵ Trafficking Protocol (n. 1), art. 3.

⁹⁶ Anne T. Gallagher, "Human Rights and Human Trafficking: Quagmire or Firm Ground? A Response to James Hathaway," *Virginia Journal of International Law* 49 (2009): 789.

⁹⁷ Janie A. Chuang, "Article 6," in Marsha Freeman, Christine Chinkin, and Beate Rudolf (eds), *Oxford Commentary on the Convention on the Elimination of All Forms of Discrimination Against Women* (Oxford: Oxford University Press, 2011).

Vigorous efforts on the part of some states and their supporters to secure an international legal prohibition on prostitution through the new definition of trafficking were ultimately unsuccessful and the drafting records affirm a right of states to deal with this matter as they see fit.⁹⁸

For much of the past decade, the US government took a different approach, deciding to use the TIP Report as a weapon in its international campaign against prostitution. It did so on the assumption that prostitution results in and fuels trafficking into sexual exploitation (widely assumed, until recently, to be the largest subcategory of trafficking),⁹⁹ and that the criminalization of sex-buying, pimping, and brothel-keeping (while decriminalizing the prostitute's act of prostitution) would lead to a reduction in sex trafficking cases.¹⁰⁰ Congress subsequently turned the latter proposition into an indicator of compliance: i.e., whether governments take measures to reduce demand for commercial sex acts and participation in international sex tourism by nationals of the country.¹⁰¹ In the past two years, there has been a dramatic shift in the TIP Report's treatment of the relationship between trafficking and prostitution, bringing the current position much more in line with accepted international standards.¹⁰² While this is a positive development, it affirms many of the concerns expressed throughout this chapter: standards are being set and shifted at the whim of the US government—to the detriment of international certainty, consistency, and unity.

Under normal circumstances, the impact of a single state's reinterpretation of a particular international legal concept will necessarily be limited. However, the newness and fragility of the international legal framework around trafficking—and the extraordinary influence wielded by the TIP Report—have removed this natural safety guard. If the Department of State decides that the cornerstone of its indicators of compliance—the concept of trafficking—is interchangeable with

⁹⁸ See, e.g., Interpretative Notes for the official records (*travaux préparatoires*) of the negotiation of the United Nations Convention against Transnational Organized Crime and the Protocols thereto, November 3, 2000, UN Doc. A/55/383/Add.1, para. 64 (affirming the individual right of states to deal with the matter as they see fit).

⁹⁹ US Department of State, Trafficking in Persons Report 2004, 15, 22 (hereinafter 2004 TIP Report) (noting the United States' "firm stand" against legalized prostitution because prostitution directly "contributes to the modern-day slave trade and is inherently demeaning").

¹⁰⁰ 2008 TIP Report (n. 85), 29.

¹⁰¹ 2005 TVPRA (n. 9) § 104(b)(1)(A).

¹⁰² The 2010 TIP Report clarifies that the US legal definitions of trafficking "are not [to be] construed to treat prostitution as a valid form of employment," while affirming that the TIP Report "evaluates the efforts of countries with legalized prostitution to reduce demand for commercial sex acts" as part of its assessment of the countries' efforts to combat trafficking. 2010 TIP Report (n. 4), 8. In a complete reversal of its previous claim that more people were trafficked into the sex sector than the non-sex sector (see e.g., 2005 TIP Report (n. 84), 6), the 2010 TIP Report confirms that "[m]ore people are trafficked for forced labor than for commercial sex," 2010 TIP Report (n. 4), 26. The report then goes on to state that a *good* trafficking law includes a definition that does not simply criminalize the recruitment or transportation of prostituted persons, and should *not* include the "related but distinct" crimes of alien smuggling or prostitution. *Ibid.*, 13. Under a section entitled "What is not Trafficking in Persons," the TIP Report lists prostitution as one category. It notes, however, that the US trafficking definition does not construe prostitution as a valid form of employment and that the TIP Report evaluates the efforts of countries with legalized prostitution to reduce demand for commercial sex acts as part of its assessment of countries' serious and sustained efforts to eliminate trafficking. *Ibid.*

forced labor and/or slavery, then it is with reference to these revised concepts that state performance will be judged. The effect is to distance the Reports, even further, from international norms. In this case, one state's view of what international law *should be* potentially becomes sufficient to displace the law *as it is*.

ii. Impact on normative development

Despite differences between US minimum standards and international law, there can be little doubt that the TIP Reports have exercised an important, sometimes decisive influence over the development and internalization of trafficking-related norms. The capacity of the Reports in this regard is at least partly an accident of history and timing. The TVPA was adopted at almost the same time as the first international legal agreement on trafficking in half a century. Between them, these two instruments occupied, for at least several years, a significant portion of the available normative space. With some important distinctions, they shared a common understanding of the nature of the problem and of the required response. This facilitated a remarkable process of normative convergence and development.

A useful example to illustrate this point relates to victim protection and support. In 2000, despite the existence of a new international treaty, the international legal obligation on states to protect and support victims of trafficking was only vaguely recognized and almost entirely devoid of substantive content. On that point, the US was slightly apart from and, many would argue, ahead of the international community. While the Trafficking Protocol carefully avoided the imposition of any hard obligation with respect to victim protection and support,¹⁰³ the TVPA had made at least some effort to develop rudimentary indicators to assess and measure state responses in this regard. Over the past decade the TIP Reports have played an important part in shaping a normative framework around this aspect of the trafficking response: moving well beyond the TVPA formal indicators in identifying "protection" as a central platform of the Reports' policy approach. There are strong indications that this approach has, in turn, fed into the international system: supporting progressive normative development through new legal instruments,¹⁰⁴ and interpretative texts attached to or supplementing older ones.¹⁰⁵

There are other examples of the TIP Reports championing a progressive standard or approach well before it received international acceptance. For example, the TVPA position that states should be judged on whether they: "ensure victims of trafficking are not inappropriately incarcerated, fined, or otherwise penalized solely for unlawful

¹⁰³ Under the Trafficking Protocol, states are only required to provide victim protection and support: "in appropriate cases and to the extent possible under domestic law." Trafficking Protocol (n. 1), arts 6–8.

¹⁰⁴ See, e.g., European Trafficking Convention (n. 3).

¹⁰⁵ See, e.g., UN Office on Drugs and Crime, Legislative Guides for the Implementation of the United Nations Convention Against Transnational Organized Crime and the Protocol Thereto, October 2004, UN Sales No. E.05.v.2, 270. For a detailed examination of the emerging obligations of protection and support, see Office of the United Nations High Commissioner for Human Rights, Commentary to the United Nations Recommended Principles and Guidelines on Human Rights and Human Trafficking (2010).

acts as a direct result of being trafficked,”¹⁰⁶ was, in 2000, a radical one. In fact, suggestions that such a standard be integrated into the Trafficking Protocol were firmly rejected by states during the negotiating process.¹⁰⁷ That US indicator, and its increasingly frequent application in the context of individual country assessments, have made a substantial contribution to transforming the non-prosecution and non-punishment principle into an accepted normative standard.¹⁰⁸ The issue of shelter detention for victims of trafficking provides another example. As little as five years ago, this practice was widespread and widely tolerated. Shortly after the release of a detailed report on shelter detention in South East Asia, the 2009 TIP Report began to question the practice.¹⁰⁹ In 2010, it presented a set of “core principles for shelter programs” in which the “ability to leave the shelter at will” was identified as essential to an effective shelter program.¹¹⁰ By 2011, states that continued to detain victims of trafficking in shelters were being singled out for individual condemnation.¹¹¹

2. Policy impacts and implications¹¹²

To what extent have the TIP indicators—and the mechanism within which they operate—changed the way in which states and the international community think about and respond to trafficking? Anecdotal evidence strongly suggests that the Reports have contributed to raising the profile of trafficking at the international, regional, and national levels; that they have nudged compliant governments to move further and faster; and that they have compelled recalcitrant ones to take steps that would otherwise have been unthinkable. Despite occasionally questioning the legitimacy of the assessments, governments worldwide appear to be very concerned over how well they rank according to US standards and about the reputational impact of their respective

¹⁰⁶ TVPA (n. 6), § 102(b)(19).

¹⁰⁷ Anne T. Gallagher, “Human Rights and the New UN Protocols on Trafficking and Migrant Smuggling: A Preliminary Analysis,” *Human Rights Quarterly* 23 (2001): 975, 990–1.

¹⁰⁸ See, e.g., UN General Assembly, “Trafficking in Women and Girls, January 30, 2009, UN Doc. A/RES/63/156, para. 12 which “[u]rges Governments to take all appropriate measures to ensure that victims of trafficking are not penalized for being trafficked and that they do not suffer from re-victimization as a result of actions taken by government authorities, and encourages Governments to prevent, within their legal framework and in accordance with national policies, victims of trafficking in persons from being prosecuted for their illegal entry or residence.” See also, European Trafficking Convention (n. 3), art. 26 (requiring states to “provide for the possibility of not imposing penalties on victims for their involvement in unlawful activities, to the extent that they have been compelled to do so”); Directive 2011/36/EU of the European Parliament and of the Council of April 5, 2011 on preventing and combating trafficking in human beings and protecting its victims, and replacing Council Framework Decision 2002/629/JHA (requiring Member States to “take the necessary measures to ensure that competent national authorities are entitled not to prosecute or impose penalties on victims of trafficking in human beings for their involvement in criminal activities which they have been compelled to commit as a direct consequence of being [trafficked]”).

¹⁰⁹ 2009 TIP Report (n. 85), 35 (“Detaining adult victim in shelters: a bad practice”), citing a report subsequently published as Anne T. Gallagher and Elaine Pearson, “The High Cost of Freedom: A Legal and Policy Analysis of Shelter Detention for Victims of Trafficking,” *Human Rights Quarterly* 32 (2010): 74.

¹¹⁰ 2009 TIP Report (n. 85), 23.

¹¹¹ See, e.g., 2011 TIP Report (n. 9), 31, 245–6.

¹¹² This section draws on and expands the analysis set out in Gallagher, “TIP Reports Review” (n. 34).

rankings.¹¹³ Most importantly in the present context, the TIP indicators themselves have been critical in securing and affirming agreement on what an effective and appropriate response to trafficking actually looks like. It is now almost universally accepted that trafficking should be criminalized; that perpetrators should be investigated, prosecuted, and punished; that trafficked persons should be provided immediate support and protection and not be re-victimized by the state; and that efforts should be made to prevent trafficking, including through addressing the many factors that increase individual and group vulnerability to exploitation.

Beyond these important but largely instinctive conclusions, the task of measuring impact, particularly in relation to individual states, is a difficult one. It is really only individual states that can accurately assess the effect of the TIP reporting process on their own behavior. However, states cannot be relied upon to provide helpful or honest insight in this regard. They have little if anything to gain by publicly admitting a link between a particular initiative or response and their assessment in the TIP Report. To date there has been very little exploration of the issue of impact and the limited research undertaken thus far provides little useful guidance on this point.¹¹⁴ There is an understandable temptation, to which the State Department itself has yielded, to use movement within the grading system as evidence of impact. A country that has advanced from Tier Three to Tier One, for example, may be identified as having responded to the pressure or threat of sanctions, as well as to the “naming and shaming” effect that accompanies highly public negative assessments of this kind.¹¹⁵ However, this interpretation fails to take account of the many political and other factors that affect a particular country’s grading in a particular year, as well as the multiple internal and external influences, beyond the Reports, on that country’s response to trafficking.¹¹⁶ In addition, there is no evidence that GTIP has recognized the growing capacity of many states to manipulate the compliance mechanism to their own ends. The authors’ own experiences in the field confirm that states are becoming increasingly adept at giving the US what its explicit and implied indicators call for, while failing to undertake the difficult, systemic changes that are really necessary to openly and effectively confront human exploitation.

Towards the future

In the field of human trafficking, the language of indicators has most frequently been employed to establish common agreement around the identifying features of a

¹¹³ Gallagher, “TIP Reports Review” (n. 34), 382, 387–9; Chuang, “Global Sheriff” (n. 48), 489–90.

¹¹⁴ See, e.g., Liana Sun Wyler and Alison Siskin, “Trafficking in Persons: U.S. Policy and Issues for Congress,” Congressional Research Service, RL34317, (2010), available at <<http://www.fpc.state.gov/documents/organization/139278.pdf>>.

¹¹⁵ H. Richard Friman, “Numbers and Certification: Assessing Foreign Compliance in Combating Narcotics and Human Trafficking,” in Peter Andreas and Kelly M. Greenhill (eds), *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict* (Ithaca and London: Cornell University Press, 2010), 75–109.

¹¹⁶ Gallagher, “TIP Reports Review” (n. 34), 387–90.

trafficking situation. The use of *indicators as indications* continues today, as the shifting, amorphous phenomenon that is human trafficking continues to resist tidy and definitive classification. However, particularly over the past half-decade, there has been increasing interest in using indicators in the more conventional sense: to measure the impact of an intervention, for example, or to evaluate compliance with a particular regime. With only a few exceptions, experiments in developing and applying indicators in these ways have not been particularly successful. The kind of standardization and precision that an indicator's approach seeks to foster does not appear to be an urgent priority for states and other stakeholders. In addition, the trafficking phenomenon itself does not appear particularly amenable to an indicator-based approach: many of the issues around trafficking are complex and contested. Key definitions and concepts are subject to multiple interpretations. The data required to accurately measure responses and impacts is often unavailable or inaccessible.

One important exception to the general reluctance to embrace an indicator-based approach is provided by the US TIP Reports. Analysis of the compliance regime established through these reports sheds new and interesting light on the challenges identified above. It also provides insight into the way in which this particular approach can shape normative development and the behavior of states. By placing itself squarely at the centre of knowledge production and global governance, the United States government has been able to play a pivotal role in global anti-trafficking efforts with its primary instrument, the TIP Reports, exercising a profound influence over the way in which states and others understand and respond to trafficking.

The TIP Reports have occupied a space that many would argue belongs properly to the international community and to the laws and structures established by that community. However, the available international machinery is weak and highly compromised. The external compliance machinery created through the TIP Reports serves an important purpose by explicitly recognizing that governments bear a responsibility to prevent trafficking and related exploitation, to end the current high levels of impunity enjoyed by traffickers, and to protect and assist victims. As one of the present authors has previously noted, attempts to persuade governments to take these obligations seriously should not fall victim to unbridled cynicism or reflexive anti-Americanism: they deserve support and encouragement.¹¹⁷

Unfortunately, the disadvantages of the current approach are substantial. The displacement of international norms through compliance indicators established and applied by one state damages international law because it is contrary to the most basic tenets of that system, including the principles of sovereignty and consent. On a more practical level, the unilateral compliance mechanism established by the TIP Reports operates to distort emerging international norms, replacing them with highly malleable standards. As demonstrated in the multiple examples explored

¹¹⁷ Gallagher, "TIP Reports Review" (n. 34).

above, the fact that states are both *understanding* and *responding* to trafficking on the basis of indicators established by a single, inconstant government, has significant implications for knowledge production as well as for global governance.

When looking to the future, it is essential to recognize that the contemporary TIP Report is operating in an environment which is very different to the one in which it was conceived. Understanding of the trafficking phenomenon has increased significantly. While controversies continue to rage, there is now a remarkable international consensus on the nature of the problem and the general direction of preferred solutions. The US monopoly on scrutiny of national responses to trafficking is being challenged, albeit somewhat feebly, by a growing body of national, regional, and international mechanisms and procedures. The international legal framework around trafficking, which, at the start of the present century was weak and ambiguous, is in much better shape. Significantly, states are becoming much more adept at delivering the minimum of what is required under the US indicators—without a noticeable change in the nature or scale of trafficking-related exploitation.

These changes will present a challenge to the US as it seeks to maintain and extend its considerable influence over how individual states, and the international community, respond to human trafficking. One possible future is suggested by recent references to the Report as “a diagnostic tool” that is neither a condemnation nor a reprieve.”¹¹⁸ Perhaps, with the experience of the past decade behind it, the virtues of nuance, modesty, and diplomacy will become more apparent, and the Report will emerge as a support and inspiration for genuinely global governance.

¹¹⁸ According to GTIP officials, the State Department’s desire to lessen the “condemnation” aspect of the report accounts for the change, in 2010, in the presentation of the TIP Report rankings. Rather than grouping countries according to tier ranking, the country ranking page was reorganized into alphabetical order—with far less shaming effect. Compare 2009 TIP Report (n. 85), 50; 2010 TIP Report (n. 4), 48; 2011 TIP Report (n. 9), 52.

Fighting Human Trafficking or Instituting Authoritarian Control?

The Political Co-optation of Human Rights Protection in Belarus

Marina Zaloznaya and John Hagan

Introduction

The idea for this chapter arose from an apparent paradox of human rights protection practices in Belarus. The Belarusian bid to join the UN Human Rights Council was rejected in 2007 due to the country's "appalling human rights record."¹ Amnesty International USA² reported regular abuses of the rights to free expression, assembly, and fair trial in Belarus; the US-based Freedom House grouped Belarus among 17 countries with "extremely oppressive environments, minimal basic rights and persistent human rights violations."³ At the same time, Belarus was lauded by the US Department of State and the United Nations Office on Drugs and Crime as a regional leader in protecting its citizens against human trafficking.⁴ Using the example of Belarus, we examine how human rights indicators can be used to create a façade behind which further human rights abuses are perpetrated. We argue that the authoritarian Belarusian government chose to pursue an eclectic combination of anti-trafficking measures with excessive zeal because they (1) improved the international image of Belarus; and (2) allowed the state to further its isolationist political agenda. We suggest that the reductionist nature of indicators creates opportunities for countries that selectively comply with

¹ Human Rights Watch. "UN 'No' to Belarus on Rights Council," May 16, 2007, available at <<http://www.hrw.org/en/news/2007/05/16/un-no-belarus-rights-council>>.

² Amnesty International USA. "Annual Report: Belarus 2011," May 28, 2011, available at <<http://www.amnestyusa.org/research/reports/annual-report-belarus-2011>>.

³ Naviny.by, "Belarus Ranks Among Worst Human Rights Abusers," April 6, 2010, available at <http://naviny.by/rubrics/english/2010/06/04/ic_news_259_332461>.

⁴ US Department of State, Trafficking in Persons Report 2010, Country Narratives A Through F, June 14, 2010, available at <<http://www.state.gov/g/tip/rls/tiprpt/2010/142759.htm>>; UN Office on Drugs and Crime, International Symposium on International Migration and Development Trafficking in Persons: Global Patterns, June 30, 2006, available at <http://157.150.195.10/esa/population/migration/turin/Turin_Statements/KANGASPUNTA.pdf>.

international human rights regulations to engage in subversive contestation of the indicator-based regime of global governance.

Repressive regimes, human rights norms, and theories of non-compliance

States' imperfect compliance with ratified international human rights agreements is well documented in academic and policy circles.⁵ Countries that are notorious abusers of human rights are surprisingly willing to sign international human rights treaties.⁶ For example, by 2006 Sudan had ratified four of seven major international human rights treaties, Algeria and the Democratic Republic of Congo had ratified six, and China had ratified five and signed two more. In comparison, the United States has ratified only three international conventions.⁷

Hafner-Burton and Tsutsui⁸ argue that the expanding international normative regime centered on human rights is a double-edged sword. On the one hand, international treaties have practically no enforcement mechanisms, offering an attractive option for "window dressing" state cooperation. Given the universal legitimacy of human rights protection, "empty" ratification is attractive to countries whose international reputations are suffering as they can diminish international pressure for reforms, even when their actual compliance is minimal.⁹ Beth Simmons calls such states *strategic ratifiers*: by endorsing international treaties, they avoid criticism, improve their international image, and develop useful diplomatic relations.¹⁰ The reduction in international pressure can then actually enable these countries to engage in increased repressive policies domestically.¹¹

In many cases treaty ratification does translate into adoption of internationally recognized standards of human rights protection, even if partial or gradual. The endorsement of human rights treaties can change the ways that governments respond to the forced disappearances of their citizens and treat their citizens during police interrogations and detention in jails and prisons.¹² However, due to lack of enforcement mechanisms embedded in international law, complete

⁵ Emilie Hafner-Burton, Kiyoteru Tsutsui, and John W. Meyer, "International Human Rights Law and the Politics of Legitimation: Repressive States and Human Rights Treaties," *International Sociology* 23 (2008): 115; Beth A. Simmons, *Mobilizing for Human Rights: International Law in Domestic Politics* (Cambridge: Cambridge University Press, 2009).

⁶ Wade Cole, "Sovereignty Relinquished? Explaining Commitment to the International Human Rights Covenants, 1966–1999," *American Sociological Review* 70 (2005): 473.

⁷ Hafner-Burton et al., "International Human Rights Law and the Politics of Legitimation" (n. 5), 119.

⁸ Emilie Hafner-Burton and Kiyoteru Tsutsui, "Human Rights in a Globalizing World: The Paradox of Empty Promises," *American Journal of Sociology* 110 (2005): 1373.

⁹ Rein Mullerson, *Human Rights Diplomacy* (New York: Routledge, 1997).

¹⁰ Simmons, *Mobilizing for Human Rights* (n. 5), 59.

¹¹ Oona Hathaway, "Do Human Rights Treaties Make a Difference?" *Yale Law Journal* 111 (2001–2002): 1935; Hafner-Burton and Tsutsui, "Human Rights in Globalizing World" (n. 8), 1398.

¹² James Ron, "Varying Methods of State Violence," *International Organization* 51 (1997): 275; Ellen Lutz and Kathryn Sikkink, "International Human Rights Law and Practice in Latin America," *International Organization* 54 (2000): 633.

non-compliance with ratified regulations is also common. Meyer et al.¹³ argue that institutional decoupling facilitates recalcitrant policies of the repressive governments. Authoritarian states tend to have low internal political integration and high autonomy that allow them to subscribe to international regulations without changing their abusive practices.¹⁴ The literature abounds with examples of repressive governments' domestic policies that blatantly contradict their international obligations.¹⁵

We argue that there is also a third model of authoritarian state behavior vis-à-vis the ratified international human rights norms whereby states use selective compliance tactics, tailored strategically to promote their own political agendas. The ratification and adoption of selective provisions allow authoritarian leaders to pursue their own interests behind a façade of compliance, while benefiting from international approval and reduced pressure for reform. In this model, selectively tight coupling rather than decoupling of regulations from on-the-ground policies enables the simultaneous compliance and worsening of authoritarian repression.

The rapidly growing norm of indicator-based international governance exacerbates the problem of selective compliance. We argue that the very nature of indicators makes them a convenient tool for countries that are abusers of human rights to improve their international image through the pursuit of select mandated policies without changing their behavior on other counts. As argued by Davis et al. in the introductory chapter of this volume, simplification and distillation of large amounts of complex, imperfect data into a condensed, often numeric, form lies at the heart of the indicator project. Through uncertainty absorption, indicators of a state's compliance generate spaces where human rights offences and other manifestations of non-compliance can be successfully hidden from the international community.¹⁶ In contrast to complete non-compliance, which rarely escapes indicators, selective compliance often allows for evasion of the equally selective assessment by the ranking entities.

In this chapter we discuss the case of Belarusian anti-trafficking initiatives as an illustration of the selective compliance model, and show how international indicators may fail to accurately assess human rights protections in cases of strategic selectivity by the ranked government. Over the past several years the international community has interpreted the ratification and strict enforcement of some anti-trafficking policies as an indicator of Belarus's growing commitment to the human rights regime. This reaction created a space within which the Belarusian government could pursue its authoritarian interests with reduced interference. Belarusian authorities used some mandated policies in ways that actually fortified the country's totalitarian regime. President Lukashenko instituted certain anti-trafficking reforms

¹³ John Meyer, and Brian Rowan, "Institutionalized Organizations: Formal Structure as Myth and Ceremony," *American Journal of Sociology* 83 (1977): 340.

¹⁴ Hafner-Burton et al., "International Human Rights Law and the Politics of Legitimation" (n. 5).

¹⁵ Linda Camp Keith, "The United Nations International Covenant on Civil and Political Rights: Does it Make a Difference in Human Rights Behavior?," *Journal of Peace Research* 36 (1999): 95; Hathaway, "Do Human Rights Treaties Make a Difference?" (n. 11).

¹⁶ Wendy Espeland and Mitchell Stevens, "A Sociology of Quantification," *European Journal of Sociology* 49 (2008): 401.

with such excessive zeal that they curbed crucial rights and freedoms of Belarusian citizens, thus increasing the physical and ideological isolation of Belarusians, augmenting the state's control over the country's institutions, and, ironically, improving the international image of Belarus.

Methodology

The data for this study were collected through the content analysis of materials prepared by international human rights organizations, statistics made available by the Belarusian government, news reports from independent and oppositional media in Belarus, materials helpfully provided by anti-trafficking NGOs in Minsk (the capital of Belarus), interviews with NGO representatives, and interviews with Belarusian students and university instructors.

First, we analyzed the accounts of the human rights situation in Belarus since the breakdown of the Soviet Union made public by the UN Office on Drugs and Crime, Amnesty International USA, United Nations Development Program in Belarus, Human Rights Watch, and the US Department of State. We then compared and contrasted these records with the information and statistics made public by (1) three Belarusian governmental or government-controlled sources: the Ministry of the Internal Affairs of Belarus (Ministerstvo Vnutrennih Del Respubliki Belarus), The National Legal Portal of Belarus (Natsional'niy Pravovoy Portal Respubliki Belarus), and the Belarusian Mission to the UN; and (2) several local independent and oppositional media and human rights organizations such as Charter97.org, Bellona.org, Belarus Digest.com, Belarusian Institute for Strategic Studies, Malady Front, Naviny.by, and others.

These data were complemented by the materials of two international anti-trafficking NGOs in Minsk—the International Organization for Migration (IOM) and La Strada. IOM is the largest inter-governmental organization addressing social issues related to migration flows worldwide, which has been operating in Minsk since 2005. La Strada, run by the YWCA and funded by several northern European governments, works specifically on the prevention of human trafficking and rehabilitation of its victims in Eastern and Central Europe since January 2003. We also conducted in-depth interviews with three employees of IOM Belarus, as well as the Director and two employees of La Strada Belarus about the work of their respective organizations, their interactions with governmental agencies, and their opinions about the state-based anti-trafficking campaign in Belarus.

Finally, we conducted a series of interviews with university students and instructors. We interviewed three political science and two international relations professors from Belarusian State University—the leading higher educational establishment in the country. We also interviewed three professors, including the co-founder and former Vice-President of the European Humanities University that was prosecuted and shut down by President Lukashenka's government. These interviews lasted one to two hours each and elicited both expert opinions

on the human rights situation in Belarus and accounts of personal experiences from the members of the Belarusian intellectual elite. We also interviewed 35 students about restrictions on their mobility resulting from the Lukashenka government's anti-trafficking campaign.

Belarus and the problem of human trafficking

According to the US Department of State's annual Trafficking in Persons (TIP) Report, the post-Soviet bloc is both the region of origin and a place of transit in human trafficking victims headed to Western Europe and the Middle East. The countries buffering Russia from the European Union offer ripe grounds for human trafficking: they are ridden with unemployment, low quality of life, and gender discrimination and violence. Every year thousands of people look for opportunities to improve their livelihood abroad, with many falling prey to human traffickers.¹⁷ Over the past six years, Belarus has putatively established itself as a regional leader in combating human trafficking. According to international indices, it has been both successful in decreasing the incidence of this crime on its own territory and in promoting various within-region and international anti-trafficking initiatives. Since 2003, Belarus has improved its ranking by the US Department of State, moving from Tier 3 to a less ominous classification as a Tier 2 country.¹⁸ The Ministry of Internal Affairs of Belarus (MIAB) reports a decline in the number of trafficking-related crimes from 555 in 2005 to 219 in 2009.¹⁹ IOM has commended Belarus for being a "vocal advocate of the need for better international cooperation in the field of anti-trafficking." The IOM reports that in the course of 2008–2009, Belarus proposed "two draft resolutions at the United Nations General Assembly calling for more rational and effective coordination mechanisms at the international level, was active in ensuring that fighting trafficking in human beings is included in the agenda of Collective Security Treaty Organization" and has continued to be "a driving force behind the drafting of the 2007–2010 CIS Programme of Cooperation on Trafficking in Persons."²⁰ In its report for the UN's Universal Periodic Review, the YWCA joined the IOM in praising Belarus for demonstrating "strong political will and undertaking significant efforts to fight human trafficking, and giving high priority to the problem nationally and internationally."²¹

¹⁷ US Department of State, Trafficking in Persons Report 2010. Country Narratives A Through F (n. 4).

¹⁸ Ibid.

¹⁹ Ministry of Foreign Affairs of the Republic of Belarus, 2012, available at <<http://www.mfa.gov.by>>.

²⁰ International Organization for Migration, "The Assessment of the National Counter-Trafficking Context and Activities in Belarus," 2006, available at <<http://www.iom.int/jahia/page809.html>>.

²¹ YWCA of Belarus, "Written Submission to Universal Periodic Review," 2010, available upon request from La Strada Belarus.

Belarus and human rights: “Appalling record”

Belarus's anti-trafficking zeal is puzzling in light of the country's continued abuse of human rights since the mid-1990s. Elected president in 1994, the authoritarian leader Alyaksandr Lukashenka has built an isolated, inward-looking state based on repeated election fraud, violent suppression of the opposition, economic dependence on Russia, and fear among its citizens.

Maintenance of a corrupt authoritarian regime in an increasingly globalized world of informational abundance requires systematic oppression of citizens' freedom of choice and expression. The US Freedom House 2010 Report ranked political freedom in Belarus with a score of 7 on a 1 to 7 scale, where higher grades indicate the lack of freedom.²² Various international observers have documented repeated abuse and voting fraud during Belarusian presidential and parliamentary elections and referenda.²³ According to Amnesty International, the Belarusian government “has cracked down on opposition leaders and movements, trade unions, non-Orthodox Christian churches, civil society organizations.”²⁴ According to the UN report, Belarus continues to use political imprisonment to silence its opposition.²⁵ Belarusian citizens run the risk of imprisonment for criticizing the government, participating in unregistered NGOs, demonstrations, and civil society organizations. Every year the US Department of State reports new cases of arbitrary detention and imprisonment without procedural rights to due process in Belarus. Numerous university students, journalists, and activists have been arrested and imprisoned for progressively oriented social action.²⁶ Anti-governmental demonstrations are routinely broken up with the use of force. Participation in unregistered organizations often results in a targeted economic pressure from the government.²⁷ There is a continuing persecution of sexual, religious, and ethnic minorities, including governmental involvement in anti-Semitic activities,²⁸ ill-treatment of

²² Charter 97, “Freedom House Called Regime in Belarus Authoritarian,” June 30, 2010, available at <<http://charter97.org/en/news/2010/6/30/30251>>.

²³ UN Human Rights, “Special Rapporteur on Belarus Condemns the Human Rights Violations during Recent Presidential Elections,” March 23, 2006, available at <<http://www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=7191&LangID=E>>; Human Rights Watch, “Belarus: Events of 2008,” 2008, available at <<http://www.hrw.org/en/node/79332>>.

²⁴ TheNewFederalist.eu, “From Freedom to Death: A Narrative of Human Rights in Belarus,” March 17, 2010, available at <<http://www.thenewfederalist.eu/From-Freedom-to-Death-A-Narrative-of-Human-Rights-in-Belarus>>.

²⁵ Charter 97, “Report on Human Rights Violations Filed to UN,” March 5, 2010, available at <<http://charter97.org/en/news/2010/3/5/27054>>.

²⁶ United States Department of State, 2008 Human Rights Report: Belarus, 2009, available at <<http://www.state.gov/g/drl/rls/hrrpt/2008/eur/119069.htm>>.

²⁷ Radio Free Europe/Radio Liberty, “Belarusian Activist Expelled from University,” May 29, 2009, available at <http://origin.rferl.org/content/Belarusian_Activist_Expelled_From_University/1742304.html>.

²⁸ NCSJ: Advocates on Behalf of Jews in Russia, Ukraine, the Baltic States, and Eurasia, “Belarus Country Page,” available at <<http://www.ncsj.org/Belarus.shtml>>.

gays and lesbians,²⁹ and state-orchestrated harassment of the Union of Poles.³⁰ Freedom of the press is also significantly limited by the state. Reporters without Borders ranked Belarus as 151st out of 175 countries in their Global Press Freedom Listing,³¹ where the countries with low rankings have the most controlled mass media. At present there are no independent TV channels, and the few existing pro-democratic Internet resources are openly persecuted by the government.³²

Why human trafficking?

Given the dubious human rights record of Lukashenka's government, its anti-trafficking crusade should be understood as an effort to consolidate authoritarianism. The beginnings of the campaign can be traced to 2004, when the international image of Belarus reached an all-time low. In April 2004, the Council of Europe condemned the human rights violations of the Belarusian government and imposed a set of restrictions on Belarusian officials' travel privileges. The US also imposed sanctions on Lukashenka's government through the Belarus Democracy Act that announced support of pro-democratic political parties, NGOs, and mass media.³³

Domestically, the situation was also precarious for the President, whose last constitutionally permitted term in office was nearing its end. Although Lukashenka was clearly unpopular, the presidential elections of 2001 resulted in his victory, with a suspicious margin of more than three-quarters of the votes. International observers challenged and the public dismissed the legitimacy of the election.³⁴ The pro-democratic Orange Revolution in neighboring Ukraine added a further threat to Lukashenka's regime. Widely supported by the West, this grass-roots protest against the falsification of presidential election results shook up the former Soviet bloc and was taken by many to signify a new era of populace-driven democratization in the region. Lukashenka responded to these internal and external pressures by increasing repressions inside Belarus while simultaneously ramping up initiatives designed to soften his reputation abroad.

²⁹ Natalya Makovska Podvergaetsya Presledovaniyu Na Rabote, September 25, 2011, available at <<http://gaybelarus.by/naviny/belarus/natalya-mankovskaya-podvergaetsya-presledovaniyu-na-rabote.html>>.

³⁰ Viasna Human Rights Center, "News on the Topic: Union of Poles," 2010, available at <<http://spring96.org/en/tags/union-of-poles>>.

³¹ Reporters Sans Frontières, "Press Freedom Index 2009," available at <<http://en.rsf.org/press-freedom-index-2009,1001.html>>.

³² Index on Censorship, "World Press Freedom Day: Belarus," May 3, 2010, available at <<http://www.indexoncensorship.org/2010/05/belarus-lukashenka-charter97-natalia-radzina>>.

³³ BBC News, "Belarus Timeline," July 10, 2010, available at <<http://news.bbc.co.uk/2/hi/europe/1118391.stm>>; Charter97, "Belarus Democracy Act Full Text," October 27, 2004, available at <<http://charter97.org/eng/news/2004/10/27/act>>.

³⁴ "Lukashenko's Rival Had No Chance," *The Moscow Times*, September 7, 2001, available at <<http://www.themoscowtimes.com/news/article/lukashenkos-rival-had-no-chance/251752.html>>; BelarusGuide.com, "Belarusian Presidential Elections 2001," October 2001, available at <http://www.belarusguide.com/politics1/Belarusian_Elections.html>.

As a part of his internal “crackdown,” Lukashenka initiated blatant repression against the strongholds of the country’s free thought—its higher educational institutions—by replacing the deans of several universities and shutting down the European Humanities University, where social sciences and humanities were taught in a liberal-arts manner.³⁵ The President also imposed stricter controls on political parties, unions, and community organizations; banned oppositional leaders from running in legislative elections; and imprisoned a prominent pro-democratic activist, Mikhail Marinich.³⁶ Simultaneously, in an effort to demonstrate popular support of his leadership to international observers, Lukashenka held an allegedly open referendum on a constitutional amendment that allowed him to run for additional terms in office, which again was reported to elicit an overwhelming support of the electorate.³⁷

In this context, the war on trafficking was a new way to assuage the West and respond to The Council of Europe’s accusations of multiple human and civil rights abuses. The choice of human trafficking was both safe and beneficial to the regime as not all human rights offenses yield themselves equally well to a plausible crusade. Given the seriousness of the problem of trafficking in the former Soviet bloc and the lack of adequate response from the regional governments,³⁸ there was little doubt that the efforts of Belarus would be noticed and commended. Most importantly, fighting against human trafficking was an attractive option because it could strengthen Lukashenka’s power by limiting the physical and ideological exposure of Belarusian citizens to different societies.

Lukashenka’s war on human trafficking

Prior to 2003, the problem of modern slavery was unrecognized officially in Belarus: there was in effect no legal basis for dealing with human trafficking crimes.³⁹ By 2007, however, Belarus became a recognized regional leader in the fight against human smuggling. In this section we will analyze the Belarusian government’s fight against trafficking and its on-the-ground implications.

In December 2004, the Interagency Commission on Crime, Corruption and Drug Trafficking of the Security Council of Belarus initiated the development of a legal and administrative framework for dealing with human trafficking and aligning

³⁵ European Humanities University, “About EHU,” 2007, available at <<http://en.ehu.lt/en/about>>.

³⁶ “Timeline Belarus,” 2010, available at <http://timelines.ws/countries/BELARUS.HTML>.

³⁷ “Belarus Referendum Results Disputed,” *Eurasia Daily Monitor*, September 2004, available at <[http://www.jamestown.org/single/?no_cache=1&tx_ttnews\[tt_news\]=27014](http://www.jamestown.org/single/?no_cache=1&tx_ttnews[tt_news]=27014)>; Mikhail Vanyashkin, “Surprise! It’s Lukashenka,” *Transitions Online* 10 (2004).

³⁸ United Nations Office on Drugs and Crime International Symposium on International Migration and Development. Trafficking in Persons: Global Patterns, June 28–30, 2006, available at <http://157.150.195.10/esa/population/migration/turin/Turin_Statements/KANGASPUNTA.pdf>; YWCA of Belarus, “Written Submission to Universal Periodic Review,” 2010, available upon request from La Strada Belarus.

³⁹ US Department of State, Trafficking in Persons Report 2002, Country Narratives A-G, available at <<http://www.state.gov/g/tip/rls/tiprpt/2002/10679.htm>>.

it with international conventions, such as the UN Convention on Transnational Organized Crime and its Protocol on Prevention and Elimination of Human Trade, especially in women and children.⁴⁰ The Belarusian government organized several conferences and training seminars on anti-trafficking regulations, bringing together governmental agencies, NGOs, international organizations, and mass media. Shortly thereafter, Belarus signed seven international conventions addressing various aspects of the fight against trafficking. Simultaneously, Belarus formed several bilateral accords with different countries regarding the law enforcement efforts in the area of human trafficking.

In September 2005, Belarus introduced a Global Partnership Initiative against Slavery and Trafficking in Human Beings in the 21st Century to the UN member states. Aimed at fostering cooperation and exchange of best practices, it led to the establishment of the Group of Friends United against Human Trafficking, consisting of 20 UN member states and the initiation of the resolution 64/178 on "Improving the Coordination of Efforts Against Trafficking in Persons," pioneered by Belarus.⁴¹ The Belarusian government also initiated anti-trafficking cooperation with the European Commission, which led to the launch of a joint project addressing the social consequences of trafficking in July 2009.⁴²

At the same time, Belarusian authorities implemented what they claimed to be preventative measures against trafficking—a "massive deliberate campaign with the utmost goal to educate Belarusians on the dangers of modern slavery."⁴³ The bulk of these initiatives were mandated by Presidential Decree # 3 "On Some Measures to Fight Human Trafficking," passed in March of 2005. The official purpose of this Act was to regulate the previously uncontrolled business and social activity believed to lead to human trafficking. First, this legislation mandated the reporting of all international travel by students to the administrations of their educational institutions, which could choose to withhold the permission to travel. Second, the decree regulated the adoption of Belarusian children by foreigners. It required approval of every adoption personally by the Minister of Education. Third, the decree required that every social medium through which people met (i.e., dating agencies), be registered with MIOB. The Decree also announced a licensing requirement for modeling agencies, mandating that all advertising materials disseminated in Belarus be produced by and feature Belarusian citizens exclusively. Arguably the most consequential part of the Decree, however, dealt with licensing of overseas employment and educational services. It required that these agencies work directly with employers/educators abroad instead of other liaisons in Belarus or abroad,

⁴⁰ United Nations Office on Drugs and Crime, *United Nations Convention against Transnational Organized Crime and Protocols Thereto*, 2004, available at <<http://www.unodc.org/documents/treaties/UNTOC/Publications/TOC%20Convention/TOCebook-e.pdf>>.

⁴¹ Belarus: Permanent Mission to the United Nations, "Human Trafficking," 2010, available at <http://www.un.int/belarus/human_trafficking.htm>; "Trafficking in Women and Girls: Meeting the Challenge Together," 2010, available at <<http://www.belembassy.org/un/conference.htm>>.

⁴² European Radio for Belarus, "EU Will Continue Supporting Belarus to Fight Trafficking in People," October 20, 2009, available at <<http://www.euroradio.fm/en/1008/reports/39721>>.

⁴³ Belarus Permanent Mission to the United States, "Human Trafficking," 2012, available at <<http://www.belembassy.org/un/human-trafficking.htm>>.

even if the latter were legally recognized and regulated by the government of the receiving country.⁴⁴ The anti-trafficking initiatives for 2008–9 were also heavily skewed toward educational outreach activities warning of the perils of foreign travel, strengthening of existing licensing requirements, controlling employment and education of Belarusian citizens abroad, prosecution of crimes, and international cooperation.⁴⁵

A closer look at Lukashenka's anti-trafficking methods reveals two core priorities. First, it is clear that international recognition for Belarus's activism was paramount for Belarusian leaders. On no other social issue has Belarus been as engaged in the international arena as it has been in the anti-trafficking fight. Belarus has become a leader among the UN member states in developing legislative and enforcement mechanisms and fostering cooperation. This role is very unusual for Belarus, given its record of extensive human rights abuses. It is apparent from the timing and substance of Lukashenka's efforts that the anti-trafficking campaign was launched to salvage the international image of Belarus and to redirect the attention of the international community away from Belarus's non-compliance on a number of other human rights regulations.

Lukashenka's domestic anti-trafficking efforts were largely oriented at the fortification of his authoritarian regime. While international norms mandate equal attention to prevention, protection of victims, and the prosecution of trafficking perpetrators,⁴⁶ Lukashenka's administration has so far been effective only on the last count. Although the state claims to excel in preventative measures, the authenticity of its efforts is doubtful. A closer look at the Belarusian anti-trafficking campaign suggests Lukashenka's measures aim primarily at restricting the physical mobility and ideological freedoms of Belarusians.

Residents of Minsk like to share a good laugh (of course, for safety reasons, in private) about the idiosyncrasies of Lukashenka, who they sarcastically call "Baŭka" ("dad" in Belarusian) with a mixture of derision and uneasiness. An urban legend has it that during one of Baŭka's walks across the capital he was appalled by the abundance of foreign models on the city's billboards. Indignant, he exclaimed: "Don't we have a country full of beautiful women? Why do I have to see the faces of Western working girls wherever I look in my country?" According to popular folklore, shortly after this incident Lukashenka introduced the restrictions on the modeling business in his 2005 Decree # 3.

Allegedly designed to keep Belarusian models from seeking employment abroad, the decree puts international companies with non-Belarusian "brand faces" in a very precarious position. The elimination of foreign business is useful in terms of the regime's commitment to decreasing its dependence on foreign manufacturing, the ideological isolation of the population, and the governmental control of the

⁴⁴ Natsional'niy Pravovoy Portal, Dekret Prezidenta Respubliki Belarus 9-go Marta 2005 g. N 3. 2005, available at <<http://pravo.by/WEBNPA/text.asp?RN=Pd0500003>>.

⁴⁵ Natsional'niy Pravovoy Portal Rspubliki Belarus, Ukaz Prezidenta Respubliki Belarus N 624. 2007, available at <<http://pravo.by/webnpa/text.asp?RN=P30700624>>.

⁴⁶ The Action Group, "Recommendations for Fighting Human Trafficking in the United States and Abroad," 2008, available at <<http://www.freetheslaves.net/Document.Doc?id=96>>.

economy. Unfortunately, licensing regulations do little to actually decrease the incidence of human trafficking. As a result of their strictness, the number of modeling agencies, and, consequently, job opportunities for Belarusian models and actors, has decreased since the passage of Decree # 3 (interview with a female anti-trafficking NGO data analyst). The two leading Belarusian modeling agencies were closed right away, while 20 smaller ones were closed in the next several months.⁴⁷ As often happens in situations when high demand is met by an artificially lowered supply, Decree # 3 provided a boost for the informal economy and, ironically, probably contributed to the proliferation of human trafficking.

Decree # 3 had a similar impact on marriage agencies in Belarus. According to the former director of the Lonely Heart agency, “there is a huge demand for this kind of service in Belarus. There are not enough men for all the women in the country; 13 percent of women in their ‘marriage age’ are lonely.”⁴⁸ While some modeling service-providers managed to get licensed under the new regulations, none of the marriage agencies obtained the new registration. Many of them transferred their efforts into the unregulated spaces of the informal economy, in all likelihood exacerbating the problem of modern-day slavery in Belarus.

Similarly, the new procedures for adoption of Belarusian orphans, mandated by the Decree, have resulted in the de facto termination of all legal adoptions by foreigners. It became practically impossible to obtain personal permission of the Minister of the Interior, for whom every adoption turned into a huge liability as the Minister risked his job if the child was trafficked. As a result, the number of adoptions by foreigners dropped from about 600 to only a few per year.⁴⁹ According to Natalya Pospelova of the Belarusian Adopting Center, with the passage of this law, “twelve thousand Belarusian orphans have been deprived of the chance to be adopted by a foreigner.”⁵⁰ At the same time, the Decree has increased the number of unregulated adoptions, significantly elevating the risks of trafficking.

Even more severe was the impact of the new licensing requirement for Belarusian liaison employment/educational agencies. The requirement for direct work with employers and education providers in the receiving country proved to be virtually impossible to satisfy. The data, collected by La Strada, show that while prior to Decree # 3 there existed more than 70 organizations that helped Belarusians study and work abroad, their number has dropped drastically in the past five years. To this day, this market niche has not returned to its previous state: as of summer 2010, there were only 50 registered agencies offering jobs abroad.

The severity of presidential regulations is illustrated by the fate of the Au Pair program in Belarus. Active in many countries worldwide for several decades, this program allows young men and women to spend up to two years abroad

⁴⁷ “Slavery and Belarus,” *Belarus News and Analysis*, April 23, 2005, available at <<http://www.data.minsk.by/belarusnews/082005/100.html>>.

⁴⁸ Ibid.

⁴⁹ Charter 97, “International League for Human Rights—Belarus Update, August 25, 2005, available at <<http://charter97.org/eng/news/2005/08/25/update>>.

⁵⁰ Slavery and Belarus, *Belarus News and Analysis*, April 23, 2005, available at <<http://www.data.minsk.by/belarusnews/082005/100.html>>.

performing baby-sitting and light household duties for their host family. The exchange is enabled through the cooperation of Au Pair agencies in the two countries, with one locating interested individuals and the other locating the families. As a rule, local governments regulate the work of Au Pair agencies; most have been in operation for a long time and have established positive reputations. According to Decree # 3, a Belarusian Au Pair agency became illegal since it was not working directly with potential employers, but, instead, relied on Au Pair agencies in hosting countries. Inasmuch as it was not feasible for the Belarusian liaison firm to directly recruit local families in Western Europe, the US, and Canada, the program was shut down. As suggested by Irina Alkhovka, the head of La Strada Belarus, “there is no doubt that licensing requirements are unjustifiably harsh. It would definitely be enough to license the liaisons that work with foreign agencies that are themselves certified by their respective governments” (June 2010, Minsk).

Decree # 3 is not effective in fighting human trafficking. According to the data collected by La Strada through the human trafficking hotline, shutting down the official liaison organizations does not decrease the number of people trying to obtain employment and education abroad. Rather, it pushes people to go through informal channels in their quest for migration. In fact, the number of hotline inquiries was the highest when the number of official agencies was the lowest.⁵¹ The new regulation has probably made foreign travel riskier for Belarusians.

Another highly consequential part of Decree # 3 mandates that educational institutions bear the responsibility for approving students’ foreign travel plans and reporting them to the Ministry of Education and the Ministry of Internal Affairs. As a result, in an effort to decrease organizational liability, Belarusian universities are discouraging their students from travelling. The account below illustrates how difficult it has become for Belarusian youth to leave the country and remain in good standing with their educational institutions.

... I was invited to a leadership workshop in Czech Republic so I needed to go for three or four days and he just wouldn’t let me go [the dean]. I was a freshman, I think, so I was way too “green” to protest. But then the next time it was different. [...] This was my senior year and I got an invitation to an academic conference in the Central European University in Budapest. [...] I felt that I could really learn something there. And all of the documents were in perfect order—I had an invitation letter from the organizers and from the hosting institution, I had a place to stay, etc. ... But, of course, I ran into a wall when I went to ask for dean’s approval. He was vehement that [...] for all he knows I could be going there “to dance on the bar tables for money.” [...] So I walked out of his office and decided that I would go anyway. When I got back, I was on the list for expulsion (27-year-old female alumna of EHU).

Many educational exchange programs were either closed or forced out of operation because selected students were not allowed out of the country. In an infamous 2005 case, 59 Belarusian students—winners of a highly selective academic

⁵¹ LaStradaBelarus, “O Programme La Strada,” 2011, available at <<http://www.lastrada.by/about>>.

competition—were denied permission to participate in an educational exchange program with the US, which had successfully operated in Belarus and other CIS countries for 13 years without a single incident of trafficking.⁵² Students' mobility is monitored closely through check-in procedures at the time of departure and return. Failure to report back to one's university upon return results in expulsion, a number of which have taken place over the past five years.⁵³

Although Belarus is vigorously promoting the UN's 3P anti-trafficking policy (Prevention, Protection, Prosecution) at the international level, the domestic strategy of Belarusian government has a law enforcement emphasis. Prevention and rehabilitation are, as a rule, delegated to the care of NGOs. Irina Alkhovka, the head of La Strada Belarus, suggests that Belarusian authorities promote deterrence through punishment because this punitive philosophy was common in the criminal justice systems of Soviet republics. Additionally, counting victims instead of preventing victimization or working with survivors has definite implications for the international visibility of Belarusian efforts. The more victims are identified, the more effective the country's anti-trafficking initiatives appear to outsiders.

The Ministry of the Interior's (MOI) investigations of trafficking crimes in Belarus are entirely based on victim testimony. Preventative investigation is extremely rare. A completed criminal offense is a precondition for action on the part of Belarusian law enforcement agencies in trafficking cases.⁵⁴ Ironically, given the extent to which the state depends on identifying victims, it consistently experiences hardships locating and talking to them. The legacy of the overbearing and repressive Soviet state, combined with the culture of fear and the authoritarian omnipotence of the current Belarusian government, generate a climate where voluntary engagement with law enforcement on the part of the citizens is extremely rare.

Belarusian trafficking victims are also reluctant to engage with the state. While human trafficking cases often involve stigmatized activities, the Belarusian MOI is not known for its delicacy and respect for confidentiality. While there is no tradition of professional ethics across any governmental social services in Belarus, Belarusian law enforcement agencies are particularly notorious for unethical and brutal treatment of citizens. Anti-trafficking NGO workers suggest that cases of unethical treatment of trafficking victims are ubiquitous. As a rule, they involve insensitivity to victims' post-traumatic disorders or breach of confidentiality. They also include instances of emotional pressure, intimidation, and arbitrary detention. The prosecution-centered activities of the government often demand forceful ways of extracting victim testimony. There are reports of the abuse of non-cooperating victims, confessions under pressure, and tweaked testimonies that allow the government to enhance the appearance of its efficiency in fighting human trafficking

⁵² Slavery and Belarus, *Belarus News and Analysis*, April 23, 2005, available at <<http://www.data.minsk.by/belarusnews/082005/100.html>>.

⁵³ Natsional'niy Pravovoy Portal, Dekret Prezidenta Respubliki Belarus 9-go Marta 2005 g. N 3. 2005, available at <<http://pravo.by/WEBNPA/text.asp?RN=Pd0500003>>.

⁵⁴ LaStradaBelarus, 2011, available at <<http://www.lastrada.by>>.

(Interviews with anti-trafficking NGO female volunteer, male volunteer, male administrative assistant, and female project manager).

According to interviews with NGO representatives, none of the victims have come to the police voluntarily. Victims prefer working with NGOs that offer social assistance and access to resources, are not associated with bureaucratic demands, and guarantee anonymity. The anti-trafficking NGOs often counsel victims in their negotiations with governmental agencies. Once victims provide their testimony—or the testimony that the governmental officials want to receive—the Ministry of Internal Affairs is no longer interested in them. Most rehabilitation work is carried out by the NGOs. La Strada, for instance, runs a rehabilitation center for trafficking survivors where the state refers victims after their repatriation.⁵⁵ According to IMO and La Strada officials, Lukashenka's administration relies on their work without providing a fair share of infrastructural and legal support (Interviews June 2010). As a result of insufficient governmental involvement, many victims of trafficking do not receive adequate psychological and physical treatment.

The efforts of MOI focus on increasing the numbers of human trafficking crimes “resolved” by the state, which offer the easiest way to persuade the international community of the effectiveness of Belarusian anti-trafficking efforts.⁵⁶ The way that the state collects statistics is widely criticized by anti-trafficking NGOs. Since the state's definitions of “crime” and “victim” always involve physical harm and material pressure, many Belarusians affected by human trafficking in less tangible ways fall outside the purview of the state (Interview with Irina Alkhovka, June 2010). Aside from several NGOs that collect their own statistics based on hotlines and direct work with clients, there are no independent data collection agencies that could evaluate the extent of human trafficking and the state's efficiency in combating it. Through licensing requirements, the government strictly controls most social and political research initiatives. Inasmuch as it is practically impossible to maintain ideological and financial independence from the state, most research organizations have either closed or moved to neighboring countries to operate in exile (i.e., Belarusian Institute for Strategic Studies and European Humanities University moved to Lithuania).⁵⁷

The strategy of the Belarusian government does not accommodate either preventative work or rehabilitation and reintegration of trafficking survivors. Instead, Belarusian efforts boil down to eclectic punitive and prohibitive measures whose effectiveness is highly debatable. The root causes, social consequences, and actual victims of modern-day slavery are not as important to Lukashenka's administration as his international reputation and preservation of authoritarianism.

⁵⁵ LaStradaBelarus, “O Programme La Strada,” 2011, available at <<http://www.lastrada.by/about>>.

⁵⁶ “Gosudarstuenniy, Uchyot Pravonarusheniyy Respublike Belarus,” 2012, available at <<http://mvd.gov.by/ru/main.aspx?gaid=1081>>.

⁵⁷ Belarusian Institute for Strategic Studies, 2010, available at <<http://www.belinstitute.eu>> European Humanities University, “About EHU,” 2007, available at <<http://en.ehu.lt/about>>.

Fighting human trafficking or maintaining the “Belarusian Miracle”?

Ideological isolation of citizens is a pillar upon which Lukashenka’s regime rests its long-term stability. The last dictatorship in Europe, Belarus thrives on its economic connections to Russia and rejects the free market foundations of the globalized economy. Both socially and economically, the country is suspended in the state of stagnation that represents Lukashenka’s idiosyncratic vision of what the Soviet state should look like in a contemporary European context. Most institutional infrastructure and the administrative tradition are preserved unaltered since the late Soviet era, with only minor concessions to the demands of rapidly developing information technology. Approximately 70 percent of the economy is state-owned; the government controls mass media, assigns people to jobs, monitors citizens’ mobility, and regulates public activity. Lukashenka’s conviction that Belarus can become a thriving economy if it follows its own development trajectory and rejects the basic lessons of globalization is known as the “Belarusian Miracle”.⁵⁸

Maintaining belief in the Belarusian Miracle within the context of an ever more interdependent world economy is progressively more challenging as information becomes accessible electronically and the Soviet past becomes more distant. The younger generation, attracted by the ideas of democracy, is much harder to convince of the virtues of totalitarianism than their parents, nostalgic for the stability of the late Soviet era. As Lukashenka is inaugurated for one presidential term after another, his administration is pressed to invent more effective methods of ideological influence. And while the pro-Belarusian Miracle propaganda is widely touted by the government, nothing works as effectively as preventing people from forming their own opinions about the regime by limiting their exposure to alternatives.

The fight against human trafficking, therefore, obscures the isolationist politics of the state. Controlling the advertising industry allows the state to be selectively prohibitive toward foreign companies believed to challenge the ideas of Belarusian supremacy. The interactions with foreigners are perceived as threatening and demoralizing for Belarusian citizens. The fight against human trafficking, then, provides a justification for limiting international interactions, whether they come in the form of an adoption, a job search, or a romantic partnership.

However, the most important tool of the state-generated hegemonic ideas is the restriction on citizens’ mobility. Belarusian anti-trafficking “preventative” measures boil down to a de facto prohibition of foreign travel. Keeping Belarusians in Belarus implies making them think Belarusian, buy Belarusian, and stay Belarusian. By strictly regulating and monitoring who is leaving the country, what they do outside

⁵⁸ “Europe’s Last Dictatorship? Just Grin and Bear It,” *TheEconomist.com*, March 20, 2006, available at <http://uniset.ca/terr/news/econ_lukashenka.html> Julia Korosteleva, “Continuity over Change: Belarus, Financial Repression, and Reintegration with Russia” in Neil Robinson (ed.), *Reforging the Weakest Link: Global Political Economy and Post-Soviet Change in Russia, Ukraine, and Belarus* (Burlington, VT: Ashgate Publishing Company, 2004), 61.

Belarusian borders, and when they come back, Lukashenka's administration is able to maintain its power within the changing geopolitical context, and with the recognition and support of the international community.

The fight against human trafficking in Belarus has assumed a peculiar moralistic tone. Due to the prohibition of travel and extensive anti-trafficking propaganda, public discourse in Belarus connects the foreign with the immoral, the different with the illicit, and the non-Belarusian with the unclean and dangerous. There are multiple examples of the moralization of foreign influences. For instance, in 2010 the Belarusian Morality Council declared that German industrial rock band Rammstein was an enemy of the state and an example of the morally degrading consequences of Western ideology.⁵⁹ Another illustration is offered by the recent attempts to introduce the so-called Economic Morality Code that would allegedly protect Belarusian citizens from the corrupt ideas of free-market proponents.⁶⁰

Another curious aspect of Belarusian anti-trafficking efforts involves the selection of its primary targets. While university students are not the main victims of human traffickers,⁶¹ the most prohibitive measures are directed primarily against them. This is not surprising given that universities are the threshold of opposition to Lukashenka's regime. The liberalization and westernization of higher education was seriously undermining his authoritarian grip as universities were encouraging analytical thinking and offering platforms for associations of students and intellectuals. By the 2001 Presidential election, student support rates for Lukashenka reached all-time lows⁶² and pro-democratic initiatives organized by students became more frequent (Interview with Vladimir Dunaev, the co-founder and former Vice-President of EHU, May 2010). It was after the 2004 pro-democratic Orange Revolution in neighboring Ukraine, where students emerged as adamant opponents of authoritarianism, that Lukashenka's state tightened its grip on universities. Several pro-Western universities were shut down, obligatory job assignments were reinstated, and students' exposure to alternative political models was limited under the guise of combating human trafficking. According to the former chair of the International Relations Department of Belarusian State University, "this decree has nothing to do with caring about the young generation. It just proves the growing tendency of the authorities to isolate the country."⁶³

Disproportionately little attention is paid to the root causes of human trafficking compared to punishing perpetrators or prohibiting citizens from leaving. Belarus suffers from many causes of human trafficking, such as high unemployment,

⁵⁹ "Belarus Declares Rammstein an Enemy of the State," *The Local*, February 23, 2010, available at <<http://www.thelocal.de/society/20100223-25443.html>>.

⁶⁰ United Democratic Forces of Belarus, "Economic Morality Code to Be Introduced in Belarus," November 23, 2009, available at <<http://udf.by/english/economy/5420-economic-morality-code-to-be-introduced-in-belarus.html>>.

⁶¹ Bureau of Justice Statistics, "Characteristics of Suspected Human Trafficking Incidents," 2007–2008, January 15, 2009, available at <<http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=550>>.

⁶² BelarusGuide.com, "Belarusian Presidential Elections 2001," October 2001, available at <http://www.belarusguide.com/politics1/Belarusian_Elections.html>.

⁶³ "Slavery and Belarus," *Belarus News and Analysis*, April 23, 2005, available at <<http://www.data.minsk.by/belarusnews/082005/100.html>>.

poverty, and ineffective social welfare. Economic problems are further exacerbated by deeply ingrained gender disparities. Belarusian women frequently bear the brunt of the economic hardship. While the number of single mothers is constantly growing, the monthly amount of governmental support is insufficient to cover the basic cost of utilities in an average apartment; women are as a rule paid less than men, kept from promotions, and fired before their male colleagues because “men need to feed the family.”⁶⁴ Although Belarusian representatives to international conventions claim that they deal with the root causes of human trafficking, the evidence from independent sources undermines these claims.⁶⁵ Twenty-eight percent of Belarusians live in poverty, the real unemployment rate is at 15 percent as opposed to one percent announced by the state, and gender discrimination and violence rates are high.⁶⁶

Domestic impact versus international reaction

The anti-trafficking policies in Belarus tend to attack the middlemen instead of dealing with the state that is the real root of the problem. An oppressive regime that does not provide space for self-realization or economic development generates powerful incentives for migration, augmenting the risks of trafficking. Yet, fighting its real causes is hardly possible without undermining the incumbent authoritarian regime. Prescribing regulations is both easy and profitable for Lukashenka’s administration. Extensive prohibitions on foreign travel and any business associated with it create the impression of governmental activity, improving the international image of Belarus, generating multiple “criminals,” and allowing Lukashenka to strengthen his authoritarian grip. The Decree # 3 contributes to the physical and ideological isolation of Belarusians, molding a docile and governable population. It allows the Belarusian government to suppress the most dangerous parts of the population—students and intellectuals—while also limiting the exposure of Belarusians to alternative models of governance, and augmenting the ideological control over all citizens by moralizing the retreat from things and thoughts foreign.

Given the actual inefficiency of Lukashenka’s domestic anti-trafficking policies, the international reaction to Belarus’s activity in this area is puzzling. In 2009

⁶⁴ Office of the High Commissioner for Human Rights, Concluding Observations of the Committee on the Elimination of Discrimination Against Women: Belarus, February 4, 2000, available at <<http://www.unhchr.ch/tbs/doc.nsf/0/84980f350885c64fc125695b00362673?Opendocument>>; United Nations in Belarus, *Belarusian Women as Seen through an Era*, 2010, available at <<http://un.by/en/publications/thema/belwomen/19-02-04-4.html>>.

⁶⁵ Statement by H.E. Natalya Petkevich, Deputy Head of the Administration of the President of the Republic of Belarus, “United Nations International Conference Trafficking in Women and Girls: Meeting the Challenge Together,” March 5, 2007, available at <http://www.un.int/belarus/conference/statement_petkevich_en.pdf>.

⁶⁶ Central Intelligence Agency, “The World Fact Book: Belarus,” 2012, available at <<https://www.cia.gov/library/publications/the-world-factbook/geos/bo.html>>; Charter97, “Real Unemployment Rate in Belarus is No Less than 15 Percent,” April 2, 2008, available at <<http://charter97.org/en/news/2008/4/2/5390>>.

Joy Ngozi, the UN Special Rapporteur on trafficking in persons, said: "I am very impressed by the political commitment demonstrated at the highest level to combat all forms of trafficking in Belarus."⁶⁷ UN press releases reported that she "commended the government of Belarus for working closely with bordering countries and other organizations involved in this issue and making significant progress. She also noted some of the good and unique practices, such as [...] extensive media campaigns directed at prevention of human trafficking."⁶⁸ Ngozi's opinion is corroborated by the Director General of the IOM, Brunson McKinley: "the government of Belarus has shown political will and determination to counter human trafficking and has achieved great results in this area."⁶⁹

Belarus and Trafficking in Persons Report: The case of subversive contestation

The approval of the international community is reflected in the US Department of State's changing attitude to Belarus. Arguably the most influential and the most trusted indicator of states' performance vis-à-vis human trafficking, the TIP Report of the US Department of State was ineffective in assessing the situation in Belarus. While no specific information provided in the report is inaccurate per se, it creates a false impression of governmental preoccupation with citizens' human rights. It does not reveal the fact that governmental activity is motivated by the interests of the regime; it does not take into consideration the de facto lack of effectiveness of Lukashenka's policies targeting low-risk populations and programs; and it does not consider strategically crafted side effects of these policies.

The TIP Report presents its assessment in two forms—through the placement of countries in one of three tiers indicating their relative effectiveness, and through brief "country narratives" that summarize the data from different sources, such as NGOs, local and Western experts, and governments themselves. Since 2003, Belarus has consistently ranked in the second tier and has received fairly positive evaluations in the country narratives.

In 2010, the narrative of the report admitted to the difficulty of assessing the situation in the country where the government is closed, independent reporting is rare, trials happen "behind closed doors," there is a lot of governmental corruption, and a general fear of the state.⁷⁰ This disclaimer, however, appears to be made for

⁶⁷ "Belarus Makes Strides in Combatting Human Trafficking, UN Expert Finds," *US News Centre*, May 26, 2009, available at <<http://www.un.org/apps/news/story.asp?NewsID=30914&Cr=human+trafficking&Cr1=>>>.

⁶⁸ Speech by Mr Antonius Broek, UN Resident Coordinator/UNDP Resident Representative about the launch of the project "Preventing, Fighting and Addressing the Social Consequences of Trafficking in Human Beings in the Republic of Belarus," January 2009, United Nations Development Programme Belarus, available at <<http://un.by/en/undp/zayavnews/belarus/21-07st.html>>.

⁶⁹ "Human Trafficking & Modern Day Slavery," 2010, available at <<http://gvnet.com/humantrafficking/Belarus.htm>>.

⁷⁰ US Department of State, *Trafficking in Persons Report 2010, Country Narratives A Through F*, 2010, available at <<http://www.state.gov/g/tip/rls/tiprpt/2010/142759.htm>>.

relieving the pressures for accuracy from the report and for transferring the responsibility of drawing conclusions about Belarus's performance from the creators of the report onto its audience. As the narrative continues with assessment and recommendations, there is also a brief mention of "reports that some policies described by the Belarusian government as anti-trafficking measures were responsible for restricting Belarusian citizens' ability to travel abroad for legitimate purposes."⁷¹ This finding is offered as a tangential side-note that has no bearing on the rest of the report. The possibility that one type of human rights protection impinges on other human rights is not deemed important enough to affect the evaluation of the country's performance.

The narrative continues with the discussion of governmental activity aimed at fighting trafficking through prosecution of offenders, prevention of offenses, and protection of victims. It suggests that Belarus is making "sustained efforts" to comply with the "minimum standards of the elimination of trafficking"; its prosecution and prevention efforts are satisfactory, while the protection of victims is worse than desired due to insufficient cooperation between the state and the NGOs and the lack of funding.⁷² Nothing in the assessment suggests that governmental anti-trafficking initiatives are actually detrimental to the basic rights of its citizens, such as the right to free expression, mobility, and political choice.

As argued by Davis et al.,⁷³ international indicators represent attempts at synthesis and simplification. The decisions regarding what is worth reporting and what can be underplayed or overlooked completely are made by a few non-experts in the name of another rationale behind the "indicators project"—standardization. Given the dearth of data on closed, non-transparent societies, the tendency toward standardization leads to further reductionism and de-contextualization.

The example of Belarus illustrates that these self-reinforcing and circular standardization-reduction processes involved in the production of indicators, make them potent tools for the reversal of power dynamics between the ranker and the ranked. Davis and colleagues suggest that the production and use of indicators create opportunities for the exercise of power through definition of standards, evaluation of compliance, and, potentially, administration of sanctions. Sometimes, however, the usual flow of power is reversed and the recipients of the assessments use their evaluations to their own advantage.

We argue that this venue for the reversal of power dynamics is particularly accessible to states that pursue selective compliance strategy. It is the inherent standardization-reduction quality of indicators that, in the case of Belarus, allowed the less powerful actor and the selective observer of human rights regulations to benefit from the ranking process. Lukashenka used the uncertainty-absorption function of indicators to bury the knowledge about the impact of his policies on

⁷¹ US Department of State, *Trafficking in Persons Report 2010, Country Narratives A Through F*, 2010, available at <<http://www.state.gov/g/tip/rls/tiprpt/2010/142759.htm>>.

⁷² *Ibid.*

⁷³ Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, "Indicators as a Technology of Global Governance," *Law and Society Review* 36:1 (2012): 71–104, esp. 73–4, 76.

other human rights of Belarusians. As the selectivity of the ranking process mapped onto the selectivity of Lukashenka's anti-trafficking policies, the resulting indicator presented an inaccurate and incomplete picture of the Belarusian situation.

The Belarusian case is an example of subversive contestation of the indicator-based regime of global governance, in which the subject of the evaluation reversed the vector of power by contesting the purpose of indicators. This contestation is subversive because, instead of openly challenging the power of the Western world over the "ranked" and the "sanctioned," the less powerful agent took advantage of the peculiarities of the ranking process itself to advance their own interests, directly contradicting the interest of the more powerful agents.

In the words of Davis and his colleagues, "indicators may mask large areas of missing or incomplete data, inability to draw significant distinctions between entities that are nonetheless hierarchically ranked, much higher levels of underlying uncertainty than the indicator depicts, and choices about weighting of different components of composite indicators." We argue that these qualities of indicators make them vulnerable to subversive contestation specifically by states pursuing selective compliance with international norms. In as much as selectivity is often hidden by reductionist indicators, the global governance regime can be contested by authoritarian states like Belarus that do not fully reject international norms but engage them selectively.

Implications for international indicators of human rights

The Belarusian case reveals the need to better understand how authoritarian states can selectively use human rights indicators to counteract the humanitarian goals of the international community. These indicators can be used to generate an unrealistically positive image of authoritarian states, while also providing a mechanism by which the international community is led to overlook abuses of human rights. While ostensibly a way of producing knowledge and encouraging reform, international indicators of human rights can actually decrease the protections available to citizens by allowing repressive governments to tighten their authoritarian controls.

The inability of the international community to correctly assess the situation in Belarus and other repressive states is due to the linear understanding of human rights and other related kinds of international indicators. The story of an authoritarian state pursuing its isolationist agenda by means of overzealously carrying out selective policies that are also associated with scoring high on international indicators of human rights protection suggests that these indicators are relational and have to be interpreted vis-à-vis each other. No state can be meaningfully judged "excellent" for selectively observing one kind of human right while simultaneously abusing many other human rights. "Human rights protection" needs to be understood as an analytical category permeated with multiple internal connections and comprised of mutually constitutive phenomena that cannot be meaningfully considered in isolation from each other. International indicators of human rights protections need to be understood and constructed in ways that reflect the

interconnectedness of human rights and the governmental activities ostensibly designed to control them. To sum up, the analysis of the Belarusian fight against human trafficking suggests that international indicators of state performance on specific issues need to be (1) broad enough to consider indirect effects of specific measures on related aspects of a country's population; and (2) also relational in further taking into account connections to other international indicators of governmental activity.

This argument is not intended to suggest that Belarus, or any other non-democratic country, should not be recognized for increased awareness of human rights abuses and compliance, even if selective and partial, with internationally recommended policies. However, as the story of Lukashenka's isolationist politics under the guise of an anti-trafficking campaign illustrates so clearly, a major problem with international indicators of human rights is not their inherent inaccuracy but, rather, their inaccurate, de-contextualized interpretation. Because it is difficult to ensure that the international community, academics, and the general public understand these indicators in light of other characteristics of individual states, the challenge is to meaningfully incorporate the context into the use of the indicators. The implications and consequences of high performance on various dimensions of human rights indicators need to be reflected in the institutional evaluations assigned to states by the indicator-producing entities. This is the only way to preclude the abuse of indicators designed by non-democratic states to cover up further infractions of human rights.

Rights-based Humanitarian Indicators in Post-earthquake Haiti*

Margaret L. Satterthwaite

Introduction

The catastrophe

The earthquake that hit Port-au-Prince on January 12, 2010 has been called “the largest urban natural catastrophe in recorded history,” killing more than 222,570 people.¹ In addition, 300,572 people were injured. Some 2.3 million people—almost 25 percent of the entire national population—was displaced.² Many of the major landmarks in the city were destroyed, and ministries and infrastructure crumbled.³ The gravity of the losses was compounded by poverty: Haiti is the poorest country in the Western Hemisphere, where more than half of the population was living on less than \$1.00 per day before the earthquake, making access to potable water and adequate food a daily struggle even before the catastrophe.⁴

* An expanded version of this chapter appeared in *N.Y.U. Journal of International Law & Politics* 43 (2011): 865. Please consult that version for complete citations and development of the arguments presented here. I am grateful to the humanitarian practitioners who shared their insights in interviews and via the online survey. I want to thank the following individuals for their insightful comments on earlier versions: Greg Beckett, Ryan Goodman, Maya Grosz, Alice M. Miller, Alison Nathan, and AnnJanette Rosga. I thank and acknowledge P. Scott Moses, who assisted with data analysis. I am grateful to Susan Hu (NYU J.D. 2011) and Valerie Brender (NYU J.D. 2012) for excellent editing and citation assistance. I thank Nora Bojar (NYU LL.M. 2010) and Colleen Duffy (NYU J.D. 2011) for research assistance. Remaining errors are solely my responsibility.

¹ “Haiti Six Months On,” UN Office of the High Commissioner for Human Rights, July 30, 2010, available at <<http://www.ohchr.org/EN/NewsEvents/Pages/Haitisixmonthson.aspx>> “Haiti: One Year Later,” UN Office for the Coordination of Humanitarian Affairs (OCHA), January 18, 2011, available at <<http://www.unocha.org/issues-in-depth/haiti-one-year-later>>.

² Ibid.

³ See “Haiti Earthquake PDNA: Assessment of Damage, Losses, General and Sectoral Needs,” Government of the Republic of Haiti, March 2010, 5.

⁴ See “Haiti at a Glance,” World Bank, August 12, 2006, available at <<http://siteresources.worldbank.org/INTHAITI/Resources/Haiti.AAG.pdf>> (comparing development indicators for Haiti and Latin American and Caribbean countries).

The world's reaction was swift. Within days, non-governmental organizations (NGOs) were providing badly needed assistance to the city's 1.5 million internally displaced persons (IDPs).⁵ In the first six months of the humanitarian response, 1.5 million people received some form of shelter; 4.3 million people were given food; and 1 million people were provided access to potable water on a daily basis.⁶ By the six-month mark, major epidemics had not appeared in the camps, and civil unrest had not materialized.⁷

Much of this work was made possible through donations to NGOs based in the United States, Canada, Australia, and Europe. Despite this generous outpouring of assistance, as of July 2010, much of the NGO money raised had not been spent, and the majority of government pledges had yet to materialize.⁸ Reports appeared decrying the lack of coordination among NGOs, between the NGO sector and the government, and between international NGOs and Haitian civil society.⁹

For displaced Haitians, life continued to be difficult and grew more perilous for many.¹⁰ As of July 2010, only 60 percent of those IDPs who were living in large camps were in locations that benefitted from coordinated servicing by the humanitarian community.¹¹ Rapes and other forms of gender-based violence were increasing, perhaps precipitously.¹² Tents and tarps, used under the baking sun for half a year in many instances, were disintegrating or tearing in the midst of the rainy season.¹³ Flooding of the camps was common and grew dangerous in some areas.¹⁴

⁵ See Humanitarian Bulletin, UN OCHA, Port-au-Prince, Haiti, June 30, 2010.

⁶ "Haiti: Six Months After," UN OCHA, July 12, 2010, 3.

⁷ Unfortunately, this did not remain the case. Cholera arrived in late 2010 and demonstrations over the international presence and the presidential elections complicated the response to the outbreak.

⁸ See Sharyl Attkisson, "Following the Aid Money to Haiti," *CBS Evening News*, May 12, 2010, <<http://www.cbsnews.com/stories/2010/05/12/eveningnews/main6477611.shtml?tag=cbsnewsTwoColUpperPromoArea>>; Joe Johns and MaryAnne Fox, "Most Countries Fail to Deliver on Haiti Aid Pledges," *CNN World*, July 15, 2010, <<http://www.cnn.com/2010/WORLD/americas/07/14/haiti.donations/index.html>>.

⁹ See "Haiti Cherie: My Dear Haiti," TransAfrica Forum, July 12, 2010, available at <<http://ijdh.org/archives/13274>>, 13 (describing lack of coordination between INGOs and Haitian civil society); "Voices from the Shanties," Inured, March 2010, http://www.inured.org/docs/Voices%20from%20the%20Shanties_INURED2010.pdf, 21.

¹⁰ For an overview of the situation in the camps as of summer 2010 based on a randomized empirical study of 1 in 8 camps, see Mark Schuller, "Unstable Foundations: Impact of NGOs on Human Rights for Port-au-Prince's Internally Displaced People," York College (CUNY)/Faculté d'Ethnologie (UEH), October 4, 2010, available at <<http://ijdh.org/archives/14855>>.

¹¹ See "Haiti 2010: Key Humanitarian Issues," Government of the Republic of Haiti, July 2, 2010, 9.

¹² See "Our Bodies Are Still Trembling: Haitian Women's Fight Against Rape," Institute for Justice and Democracy in Haiti, Madre, TransAfrica Forum, Univiversity of Minnesota Law School Human Rights Litigation and Advocacy Clinic, and University of Virginia School of Law International Human Rights Law Clinic, July 2010, available at <http://reliefweb.int/sites/reliefweb.int/files/resources/2AFAD9E18B0B66604925776E000646D1-Full_Report.pdf>; see also "Adolescent Sexual and Reproductive Health Assessment Mission to PLAN Haiti," InterAct Worldwide, July 2010, 5.

¹³ OCHR estimated that 40 percent of the tents and many of the tarps distributed since the earthquake needed to be replaced by July 2010. "Haiti: Six Months After," UN OCHA, July 12, 2010, 3.

¹⁴ See *ibid.*, 2.

Forced displacements had begun, spurred on by private landowners seeking to clear their land of IDP tent camps.¹⁵

Without adequate mechanisms to hold NGOs to account, expressions of dissatisfaction took different forms. In the rainy heat of Port-au-Prince, graffiti reading “*Aba ONG vole*” (Down with NGO thieves) began to appear on walls throughout the city.¹⁶ Several aid workers were kidnapped.¹⁷ Foreign humanitarians huddled in their compounds in the evenings, afraid of becoming targets of IDP frustration.

In this fraught context, the desire to measure progress is compelling. Very often the assessment of progress in the humanitarian sphere has involved the prominent use of indicators: project indicators, sector indicators, and industry-wide indicators. These metrics are at the heart of the humanitarian endeavor today, including in post-earthquake Haiti. But these metrics, aimed at responding to demands for quality and accountability, raise many questions of their own. For example, who is the audience for such indicators? Does striving to achieve universal indicators frustrate efforts to respond to local needs? This chapter sets out to examine these questions, drawing on empirical and critical methods, and focusing on the use of indicators in Haiti.

Using Haiti as a case study, this chapter examines one leading set of standards and indicators developed by professional humanitarians in the past dozen years that have as their aim improving the quality, effectiveness, and accountability of their own response to disaster. In order to create a finite period for analysis, this chapter examines only the first six months after the January 12, 2010 earthquake. The chapter builds on analysis of human rights indicators previously carried out with AnnJannette Rosga.¹⁸ Understanding industry-wide humanitarian indicators as a “technology of global governance,” the chapter also draws on the theoretical framework set out by Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry in their work on indicators.¹⁹

The chapter proceeds as follows: this section will conclude by presenting the chapter’s methodology and the key findings of an online survey of humanitarians undertaken for this study. The next section presents a brief overview of the Sphere Project’s Minimum Standards in Disaster Response (“Sphere” or “the Sphere Project”) and then deploys a governmentality analysis to suggest that sites of emergency are viewed as “calculable and manageable spaces” through the diffusion

¹⁵ See Alice Speri, “Still Homeless from Haiti Earthquake, Thousands Fight Forced Evictions,” *Christian Science Monitor*, July 2, 2010, available at <<http://www.csmonitor.com/World/Americas/2010/0702/Still-homeless-from-Haiti-earthquake-thousands-fight-forced-evictions>> (noting that 206 refugee camps among 1,241 were recognized by UN OCHA as of July 2010).

¹⁶ Notes from author site visit to Port-au-Prince (June 2010) (on file with author).

¹⁷ See Marine Hass, “Two Kidnapped European Aid Workers Freed in Haiti,” *Reuters*, March 11, 2010, available at <<http://www.alertnet.org/thenews/newsdesk/N11251709.htm>>.

¹⁸ See AnnJannette Rosga and Margaret Satterthwaite, “The Trust in Indicators: Measuring Human Rights,” *Berkeley Journal of International Law* 27 (2009): 253.

¹⁹ See Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, “Indicators as a Technology of Global Governance,” *Law and Society Review* 36:1 (2012): 71–104. This chapter considers the Sphere indicators to be amenable to the type of analysis suggested by Davis, Kingsbury, and Merry, although they are not rank-ordered and are not always numerical. The Sphere indicators otherwise match the definition developed by Davis et al., *ibid*.

and use of professional standards and indicators.²⁰ Humanitarian indicators allow for governance at a distance, drawing on the logics of audit, quantification, and management, as well as the language of human rights.²¹ With this framework as a starting point, the following section presents the results of empirical analysis of indicators and benchmarks in the post-earthquake humanitarian crisis in Haiti. Data collected through an online survey, site visits, and interviews with human rights experts suggest that humanitarians are using indicators and benchmarks to monitor projects and ensure effective, measurable results in Haiti. Such metrics also have unintended impacts in Haiti. While humanitarians have managed to deliver impressive, life-saving services to IDP camps in incredibly difficult conditions, the system has also failed to ensure predictable services to large swathes of the population. In part, this may be because issues of coverage and scope are not as visible via monitoring and assessment tools such as indicators, as are project quality and outcomes. Although some indicators include attention to issues of sustainability, by highlighting successes in improving outcomes, they also inadvertently downplay the potential damage that humanitarian interventions can do to existing and nascent systems for delivering key services. Finally, the quest for data seems to become less pressing when protection issues arise: the humanitarian system appears reluctant to count rapes and evictions despite rising calls to monitor such abuses.

Methods

This chapter combines critical analysis with empirical findings drawn from author interviews, site visits, and an online survey, and seeks to examine the use of indicators and benchmarks through a governmentality framework.²² The author conducted two dozen semi-structured interviews between May 23 and July 7, 2010. About a dozen of the experts interviewed were humanitarians engaged in the design, drafting, and revision of the various editions of the Sphere Project's Minimum Standards in Disaster Response ("Sphere" or "the Sphere Project"), the Humanitarian Accountability Partnership's Humanitarian Accountability Standard ("the HAP Standard"), and several other standards and indicators projects. To supplement these views, the author conducted interviews with human rights experts familiar with humanitarian standards and indicators, and those with Haiti-specific experience. In addition, the author visited Haiti twice during the research for this chapter: once in March 2010 and once in June 2010. While there, the author witnessed earthquake damage, viewed IDP camps, attended humanitarian

²⁰ Nikolas Rose, Pat O'Malley, and Mariana Valverde, "Governmentality," *Annual Review of Law and Social Science* 2 (2006): 83.

²¹ *Ibid.*

²² The concept of "governmentality" is drawn from Foucault. See Michel Foucault, in Michel Senellart (ed.), *Security, Territory, Population: Lectures at the Collège de France 1977–1978* (New York: Palgrave Macmillan, 2007). For an overview of this scholarship, see Rose et al., "Governmentality" (n. 20). This approach uses the concept of the mentality of governance—"governmentality"—to interrogate forms of contemporary rule that extend beyond the state.

Cluster meetings in Port-au-Prince, and met with humanitarians and human rights advocates. Data gathered during these visits is integrated throughout the chapter.

To understand the use of indicators by humanitarian workers in Haiti, the author conducted an online survey in English and French between May 18 and June 15, 2010. One hundred and thirty-eight unique respondents—individuals who met the inclusion criteria of working in the humanitarian or development field and having worked full-time in or in relation to Haiti at some point during the preceding two years—filled out the survey. The survey results suggest that indicators, in general, are routinely used among humanitarians in Haiti,²³ and that project-based indicators were used with some frequency. Such standards are often based on existing industry standards like Sphere, which was familiar to respondents (76 percent reported being familiar with Sphere). This means that although the proportion of respondents using Sphere may seem relatively small, the influence of such standards and indicators extends beyond their direct use. When asked about the impact of using indicators and benchmarks, humanitarians overwhelmingly reported that such use strengthened accountability to both beneficiaries and donors, ensured alignment with human rights standards, enhanced inclusion and participation of beneficiaries, increased transparency to stakeholders, and improved both the efficiency and the impact of their work. While the research for this chapter was not designed to determine whether humanitarian indicators are in fact improving quality, efficiency, and accountability, the results of the survey and interviews suggest that humanitarians overwhelmingly believe that indicators have such impacts. This belief alone means that indicators are worthy of careful attention as a technology of global governance.

Unpacking the logics of humanitarian indicators

The development of the sphere project

The formal humanitarian aid system is made up of a wide array of actors who have as their purpose assisting disaster-affected communities in their quest for survival and dignity. Alongside the International Federation of the Red Cross and Red Crescent Societies (IFRC), are a variety of international NGOs (INGOs), key UN agencies concerned with crisis such as the Office for the Coordination of Humanitarian Affairs (OCHA), the Department of Peacekeeping Operations, and the emergency departments of the UN Children's Fund (UNICEF) and UN Population Fund (UNFPA),

²³ Seventy percent of the survey respondents reported that they use indicators in their Haiti-related work; of these, 74 percent use them at least weekly and a full 94 percent use them at least monthly. A striking 85 percent of those responding reported using a rights-based approach in Haiti. Seventy-three percent of those who answered a question about which type of standards they use reported using human rights treaties in relation to their work in Haiti; 62 percent use guidelines or standards prepared by the humanitarian system's Inter-Agency Standing Group; 45 percent use the HAP Standard, and 33 percent use Sphere. Respondents reported using indicators and benchmarks for a wide variety of activities.

and donor governments.²⁴ This formal system has developed institutions, networks, and practices that stretch across the globe, springing into action with set coordination structures, joint funding mechanisms, and common understandings of best practices. The formal system operates alongside an “informal” humanitarian system that includes affected governments, militaries, and local communities. The informal system is routinely part of crisis response but its composition, relationship to the formal system, and degree of participation in the various structures created by the international humanitarian community is context-dependent and widely variable.

In 2007 and 2008 alone, the formal humanitarian system responded to 52 humanitarian emergencies in the world.²⁵ In 2008, there were an estimated 26 million internally displaced persons and 11.4 million refugees in the world,²⁶ and approximately 210,800 emergency and rehabilitation workers responding to crises.²⁷ Although about 95 percent of INGO staff is drawn from the local population, the direction and policies of those organizations are, for the most part, dictated from headquarters offices based in the Global North.²⁸ The six largest INGO federations had a combined expenditure of over \$4 billion, with \$1.7 billion allocated specifically to disaster programming.²⁹ Just over half of these resources come from private (individual, foundation, and corporate) sources, with the remainder provided by multilateral and state donors.³⁰ The broader crisis industry, which includes for-profit enterprises, has been estimated to command \$10 billion per year.³¹

To regulate and improve the performance of the massive humanitarian aid system and to respond to some of its earlier failures, humanitarian NGOs set up the Active Learning Network for Accountability and Performance (ALNAP) and the “Standards Project” in 1997, aimed at formulating standards for humanitarian assistance. The Standards Project became the Sphere Project, which seeks to “improve the quality and accountability of performance by humanitarian professionals” through the development of substantive standards and measurable indicators.³² By 2000, Sphere had created a set of standards, indicators, and guidance

²⁴ See Paul Harvey, Abby Stoddard, Adele Harmer, and Glyn Taylor, “The State of the Humanitarian System: Assessing Performance and Progress—A Pilot Study,” ALNAP, January 2010, available at <<http://www.alnap.org/pool/files/alnap-sohs-final.pdf>>, 14 for a discussion of the formal and informal systems.

²⁵ Ibid.

²⁶ “Humanitarian Accountability Report 2009,” HAP International, May 2010, available at <<http://www.hapinternational.org/pool/files/b-the-2009-hap-secretariat-report.pdf>>, 12 (hereinafter HAP International, Humanitarian Accountability Report).

²⁷ Harvey, et al., “The State of the Humanitarian System” (n. 24), 18.

²⁸ Ibid., 18–20.

²⁹ Ibid., 20.

³⁰ HAP International Humanitarian Accountability Report, May 2010, 17.

³¹ Craig Calhoun, “The Imperative to Reduce Suffering: Charity, Progress, and Emergencies in the Field of Humanitarian Action,” in Michael Barnett and Thomas C. Weiss (eds), *Humanitarianism in Question: Politics, Power, Ethics* (Ithaca and London: Cornell University Press, 2008), 89 (hereinafter *Humanitarianism in Question*).

³² “Quality and Accountability Initiatives: Questions and Answers,” ALNAP, HAP, People In Aid, and Sphere Project, October 2006, available in Spanish at <http://www.alnap.org/pool/files/QA_Spanish.pdf>. It is in English at <<http://www.hapinternational.org/pool/files/q-&-a.pdf>>.

notes through an extensive consultative process.³³ The standards are organized by sector: water supply and sanitation; nutrition; food aid; shelter and site planning; and health services.³⁴ The first print version of the standards was published in the *Sphere Handbook* in 2000; this book soon became ubiquitous in field settings. To ensure that the standards and indicators remained state-of-the-art, a fully revised version was published in 2004, and a third edition was published in 2011.

The standards and indicators are explicitly rights-based: reference is made to the rights framework in the technical chapters and a set of “cross-cutting issues” is integrated throughout the *Handbook*. Gender equality, children’s rights, and protection are incorporated into the technical chapters and their rights basis is explained in a brief introduction presenting the cross-cutting themes as “important issues that have relevance to all sectors.”³⁵ Common standards address: participation; initial assessment; response; targeting; monitoring; evaluation; aid worker competencies and responsibilities; and supervision, management, and support of personnel.³⁶ These cross-cutting issues and the “Common Standards” chapter set out principles that are said to be drawn from human rights.³⁷ The rights-based language in *Sphere* has been deepened and expanded in the 2011 edition of the *Handbook*.³⁸ That edition will not be exhaustively analyzed, since the 2004 edition was in use in Haiti during the period under analysis.

Humanitarians governing at a distance via indicators

Using rights-based approaches, standards, and indicators, professional humanitarian actors govern the putatively ungoverned—those whose governments have collapsed, become predatory, or can no longer be counted on to provide protection.³⁹ At a material level, in many emergency settings, major INGOs effectively take the place of

³³ *Ibid.*, 2.

³⁴ “Humanitarian Charter and Minimum Standards in Disaster Response,” Sphere Project, 2000 edition 1 (hereinafter Sphere Project 2000 ed.).

³⁵ *Ibid.*, 10–13.

³⁶ *Ibid.*, 24.

³⁷ The Common Standards chapter explains the standards’ links to human rights as follows. Humanitarian agencies have the responsibility to provide assistance in a manner that is consistent with human rights, including the right to participation, non-discrimination and information, as reflected in the body of international human rights, humanitarian and refugee law. In the Humanitarian Charter and the *Code of Conduct for the International Red Cross and Red Crescent Movement and Non-Governmental Organizations in Disaster Relief*, humanitarian agencies undertake to make themselves accountable to those they seek to assist. The common standards outline the responsibilities of organisations and individuals when providing protection and assistance.

“Humanitarian Charter and Minimum Standards in Disaster Response,” Sphere Project, 2004 edition, available at <<http://ocw.jhsph.edu/courses/refugeehealthcare/PDFs/SphereProjectHandbook.pdf>>, 10–13 (hereinafter Sphere Project 2004 ed.).

³⁸ See, e.g., the new chapter on “protection principles” in the revised 2011 edition. “Humanitarian Charter and Minimum Standards in Disaster Response,” Sphere Project, 2011 edition 25–47.

³⁹ Duffield suggests that emergency actors govern when the ungoverned become unmanageable or uncontained. Mark Duffield, *Development, Security and Unending War: Governing the World of Peoples* (Cambridge: Polity Press, 2007), 111–32.

the state—or large swathes of it—for disaster-affected populations. They provide or facilitate access to the basic material goods needed for human survival, such as food, water, and shelter, and seek to impose order from chaos by setting up rational, managed camps and other systems where people's needs are satisfied. They also seek—although to a lesser extent than in relation to material goods—to protect disaster-affected populations from certain types of security threats, such as sexual violence and child exploitation. These INGOs command enormous resources—in many cases, more than the government ministries whose tasks they often take on. To give a sense of the scale of humanitarian INGO resources: in 2007, the French NGO Médecins Sans Frontières spent more on humanitarian programming than any single Western state donor except the United States and the European Union.⁴⁰

The growth of the INGO sector has been analyzed as—among other things—a complex form of outsourcing by Western and Northern donor states. Where Western states governed in the past through direct rule via colonialism and indirect rule via Cold War-era neocolonialism, such powers now outsource significant portions of their governance activities in the Global South to humanitarian and development INGOs. Mark Duffield explains that in the early 1980s, the majority of humanitarian aid was given directly to affected country governments via their relevant ministries.⁴¹ This began to shift, however, as concerns about Southern state corruption grew. Donors began to increase their humanitarian aid to INGOs, which asserted an “organizational neutrality . . . between Third World Corruption and Western complicity.”⁴² Duffield argues that in asserting their independence from any government—donor or host—INGOs had a comparative advantage since they could promote Western interests from a formally independent position while also working closely with Southern governments and populations.⁴³ With significant material resources and a specific form of “organizational neutrality,” INGOs developed a type of “sovereignty” by the late 1990s.⁴⁴ This INGO sovereignty was made possible through another shift in funding: since the late 1990s, the proportion of overseas governmental aid devoted to development and emergency programming shifted proportionally, with dramatic increases in emergency aid and concomitant decreases in development aid.⁴⁵

Commanding significant resources, INGOs do not only provide subsistence goods and services. They are also engaged in more obvious governance efforts. In exercising their recognized power over the populations they serve, they set up camp management structures charged with ensuring fair access to goods and services, consult with camp committees and advisory groups, and ensure that vulnerable members of communities have priority access to services.⁴⁶ They conduct census

⁴⁰ HAP International Humanitarian Accountability Report, May 2010, 17.

⁴¹ Duffield, *Development, Security and Unending War* (n. 39), 72.

⁴² *Ibid.*

⁴³ *Ibid.*, 90–1.

⁴⁴ See generally, *ibid.*, 72, 91 (developing the concept of INGO “petty sovereignty”).

⁴⁵ See James D. Fearon, “The Rise of Emergency Relief Aid,” in Barnett and Weiss (eds), *Humanitarianism in Question: Politics, Power, Ethics*, 49–51.

⁴⁶ See, e.g., “Camp Management Terms of Reference,” IASC CCCM Cluster, on file with author.

counts, keep data on who enters and departs from camps, and work to understand the needs and vulnerabilities of those under their care. Alex de Waal points out that these “emergency” actions can last many years—even decades.⁴⁷

Social science scholars who use the concept of “governmentality” to interrogate forms of rule in current times that extend beyond the state provide helpful insights into the tools of governance utilized by humanitarian INGOs.⁴⁸ Governmentality analysis is especially fitting in relation to INGOs, which Duffield suggests have been “governmentalized.”⁴⁹ Michel-Rolph Trouillot has theorized that NGOs produce wide-ranging “state effects” through planning, interpellation, and legibility practices in the Global South, giving Haiti as one example.⁵⁰ In relation to Haiti, Erica Caple James has explained, “In the case of transnational humanitarian assistance the practices of governmentality often occur in place of the weak or failed state.”⁵¹

One of the forms of dispersed rule that governmentality scholars have identified is what Rose and Miller call “governing at a distance.”⁵² This is a dynamic through which agents of governance, “acting from a center of calculation such as a government office or the headquarters of a nongovernmental organization,” impact the “activities of others who were spatially and organizationally distinct.”⁵³ This concept is very helpful in analyzing the dynamics of governance through indicators in the humanitarian realm.

There are three relevant dynamics that might be fruitfully understood as governing at a distance in the emergency context. The first and most obvious is the relationship already discussed, whereby INGOs govern disaster-affected populations through the provision of certain types of subsistence goods and services, and through management techniques such as census and the creation of camp management committees. The second is the relationship between Northern-based INGO direction and Southern-based INGO line managers and field staff (as well as local NGO implementing partners). While those “in the field” are involved in the day-to-day dynamics of governing disaster-affected populations, the upper management is setting the policies, frameworks, and organizational structures through which

⁴⁷ Alex de Waal, “The Humanitarians’ Tragedy: Escapable and Inescapable Cruelties,” *Disasters* 34 (2010): S130, S136 (citations omitted).

⁴⁸ This scholarship is based on Foucault’s later writings. See Michel Foucault in Michel Senellart (ed.), *Security, Territory, Population: Lectures at the Collège de France 1977–1978* (n. 22). For an overview of governmentality scholarship, see Rose et al., “Governmentality” (n. 20). Rose, O’Malley, and Valverde explain that governmentality scholars ask a set of questions about the techniques, objects, and ends of governing that are tailored to analyze the technologies used by different governing bodies throughout a dispersed system of power (“Governmentality” (n. 20)).

⁴⁹ See generally Duffield, *Development, Security and Unending War* (n. 39), 72.

⁵⁰ See Michel-Rolph Trouillot, “The Anthropology of the State in the Age of Globalization,” *Current Anthropology* 42 (2001): 125, 132 n. 22. Mark Schuller echoes this: Mark Schuller, “Seeing Like a ‘Failed’ NGO: Globalization’s Impacts on State and Civil Society in Haiti,” *Politics and Legal Anthropology Review* 30 (2007): 67.

⁵¹ Erica Caple James, “The Political Economy of ‘Trauma’ in Haiti in the Democratic Era of Insecurity,” *Culture, Medicine and Psychiatry* 28 (2004): 127, 131.

⁵² See Peter Miller and Nikolas Rose, *Governing the Present* (Cambridge: Polity Press, 2008): 22, 33–4.

⁵³ Rose et al., “Governmentality” (n. 20).

that governance takes place. Major policy decisions, such as whether an INGO will use the Sphere standards and indicators, are generally made by headquarters staff. The third relationship is between donor states and INGOs, a relationship that is especially helpfully understood through the concept of governance at a distance. In part this governance occurs though forbearance: both donor states and INGOs are most effective when INGOs are understood as a formal matter to act independently, free of coercive legal requirements or strong-arming. Instead of such coercive practices, donors govern at a distance through the bureaucratic and technical exercises associated with granting funds, planning programs, and conducting monitoring and evaluation of “outcomes” and “accountability.”⁵⁴

This language of outcomes linked with accountability sounds in the specific discourse of project management utilized by Western donors and INGOs. These forms, known as “results-based management” (RBM) and the “logical framework” or “logframe,” came into vogue in the INGO world as state donor agencies adopted them widely in the 1980s.⁵⁵ As INGOs began to receive more and more official development aid for emergency programming, they adopted the RBM and logframe approaches alongside their donors in order to demonstrate their accountability to the outcomes for which they were receiving funding. These frameworks use indicators as a key element.⁵⁶ The logics behind logframes and RBM were proliferating within the humanitarian system by the time Sphere was under development. It is no wonder, then, that the same logic underpinning these rational management tools became the logic undergirding the Sphere standards and indicators.

Humanitarian indicators in Haiti

Historical context

It is impossible to discuss Haiti sensibly without referring to the country’s fraught history. The founding of the nation, its troubled political history, its position in the

⁵⁴ Janice Gross Stein, “Humanitarian Organizations: Accountable—Why, to Whom, for What, and How?” in Barnett and Weiss (eds), *Humanitarianism in Question*, 126, 127 (describing the increased field presence of NGOs and decreasing presence of states, which “contract out” their responsibilities to NGOs and exert influence through monitoring and evaluation instead of direct engagement).

⁵⁵ Tina Wallace with Lisa Bornstein and Jennifer Chapman, “The Management of Development,” in *The Aid Chain: Coercion and Commitment in Development NGOs* (London: Intermediate Technology Publications Ltd, 2006), 31, 32–3. For a discussion of the adoption and use of logframe analysis within the humanitarian INGO context, see Monika Krause, “The Logic of Relief: Humanitarian NGOs and Global Governance,” 89–93 (unpublished PhD dissertation, New York University, 2009) (on file with author).

⁵⁶ *Ibid.*, 34 (“The logframe is the most well-known tool for aid disbursement, used by almost all donors and NGOs in the UK, Europe, and US; it is the basic project document that includes project goals, plans, timetables for implementation, required inputs and expected outcomes with associated measurements (indicators), and an analysis or listing of external factors and internal assumptions that may be a risk to the achievement of the goals.”)

global economic order and political imaginary, and its recent experiences of crisis and intervention, are all relevant to the current situation in Haiti. Briefly:

In August 1791, slaves in northern Saint-Domingue launched an uprising that spread throughout the colony and turned into a successful revolution that toppled both slavery and the French colonial order. The revolution took nearly thirteen years to unfold from the initial uprising to the proclamation of Haitian independence in January 1804.⁵⁷

This proclamation was followed by a vigorous containment strategy imposed by France and its allies. Haiti was coerced into paying an indemnity to France in exchange for the normalization of economic and diplomatic relations.⁵⁸ The indemnity payment was enormous, amounting to double the price paid by the United States for the territory gained in the Louisiana Purchase.⁵⁹ This payment is credited with directly contributing to Haiti's current situation.⁶⁰

Haiti's recent history has been one of "permanent crisis," according to anthropologist Gregory Beckett.⁶¹ Since 1994, when the US military restored then-president Jean-Bertrand Aristide to power by ousting the ruling *de facto* regime, international engagement in Haiti has been conducted through the concepts and frameworks of state failure and cyclical emergency.⁶² These frameworks shorten planning horizons and prioritize quick action over long-term change or justice.⁶³ This has meant, in material terms, that aid flows which would otherwise be targeted to building, improving, and expanding Haiti's public infrastructure and systems, have in large part been channeled to INGOs, Haitian NGOs, and other donor-influenced entities charged with implementing programs.⁶⁴ As a result, the state—already weak and ineffective—was in important ways significantly *de-capacitated* through foreign aid instead of being strengthened.⁶⁵ This meant not only that governance was weakened, but also that systems—such as those providing healthcare, education, and potable water—were likewise starved of resources. To be sure, INGOs and some Haitian NGOs stepped in to provide aid in many cases, but

⁵⁷ See Michel-Rolph Trouillot, *Silencing the Past: Power and the Production of History* (Boston: Beacon Press, 1995), 37.

⁵⁸ Laurent Dubois, *Avengers of the New World: The Story of the Haitian Revolution* (Cambridge, MA and London: Harvard University Press, 2004), 303–4.

⁵⁹ Brian Concannon, Jr and Anthony Philips, "Haiti Needs Justice, Not Charity," *South Florida Sun-Sentinel*, July 26, 2006, available at <<http://www.globalpolicy.org/component/content/article/186/34476.html>>.

⁶⁰ *Ibid.*

⁶¹ Gregory A. Beckett, "The End of Haiti: History Under Conditions of Impossibility," 5 (unpublished PhD dissertation, University of Chicago, 2008) (on file with author).

⁶² *Ibid.* See also Greg Beckett, "Phantom Power: Notes on Provisionality in Haiti," in John D. Kelly, Beatrice Jauregui, Sean T. Mitchell, and Jeremy Walton (eds), *Anthropology and Global Counterinsurgency* (Chicago: University of Chicago Press, 2010) (analyzing the operation of the "twin concepts of state failure and the state of exception" in relation to the forced removal from power of President Jean-Bertrand Aristide in 2004).

⁶³ See Erica Caple James, *Democratic Insecurities: Violence, Trauma, and Intervention in Haiti* (Berkeley: University of California Press, 2010), 178–222.

⁶⁴ See Terry F. Buss with Adam Gardner, *Haiti in the Balance: Why Foreign Aid has Failed and What We Can Do About It* (Washington: Brookings Institution Press, 2008), 118–21.

⁶⁵ Schuller, "Seeing Like a 'Failed' NGO" (n. 50), 73 notes that as foreign aid was directed to NGOs, the Haitian state became an "apparent state" (internal citations omitted).

the impact of this dynamic was that the resulting systems themselves did not belong to the state, were in important respects unregulated, and could not be counted on to sustainably provide services to all in an accountable manner.⁶⁶ By the time of the earthquake, Haiti was ranked number twelve out of 177 states in the Fund for Peace's Failed States Index (where 1 represents greatest failure), and 129th of 141 states in the Index of State Weakness in the Developing World (where 1 represents the strongest state).⁶⁷

When the earthquake hit, it decimated the already weakened public sector in Haiti: 27 of 28 national government buildings were damaged, and somewhere near 20 percent of the whole Haitian civil service were killed or injured.⁶⁸ The massive humanitarian effort, so badly needed, unsurprisingly bypassed the badly shocked Haitian government in the first few weeks of the crisis. By the end of January 2010, less than 1 percent of the earthquake aid had gone to the government of Haiti.⁶⁹ As understandable as it may have been in those first few weeks, this dynamic was slow to change. As of July 2010, UN Deputy Special Envoy to Haiti, Dr Paul Farmer reported that "of \$1.8 billion for earthquake relief sent to Haiti, less than 2.9 percent has so far gone to the government."⁷⁰ In a setting like this, what impact do rights-based indicators have? Does their deployment bring these human rights issues into view, or obscure these dynamics even further?

Spontaneous settlements and unmanaged camps

In the aftermath of the January 12 earthquake, surviving members of poor Haitian families whose houses had been destroyed took what belongings they could salvage from their homes and fled to the closest or most familiar open spaces. Flocking to public squares, parks, schoolyards, and golf courses, Haitians set up makeshift shelters using whatever materials they could find. In Port-au-Prince, the humanitarian system responded by setting up camp management agencies (CMAs) within the spontaneous settlements. These CMAs were charged with identifying the needs of residents, coordinating and prioritizing aid, ensuring adequate governance systems within the camps, and assessing the camp sites for risk of landslide and flooding.⁷¹

⁶⁶ For a discussion of this dynamic in Port-de-Paix, northern Haiti, see Partners In Health, CHRCJ, RFK Memorial Center for Human Rights and Zanmi Lasante, "Wòch nan Soley: The Denial of the Right to Water in Haiti," 2008, available at <<http://www.chrcj.org/projects/docs/wachmonsoley.pdf>>.

⁶⁷ See Rencoret et al., *Haiti Earthquake Response: Context Analysis* (ALNAT & UNEG, July 2010), 8.

⁶⁸ Testimony of Dr Paul Farmer, July 27, 2010, <<http://www.pih.org/haiti/news-entry/focus-on-haiti-the-road-to-recovery-a-six-month-review/#farmer>> (hereinafter *Farmer Testimony*).

⁶⁹ See *Haiti Government Gets 1 Cent of U.S. Aid Dollar*, MSNBC.COM, January 27, 2010, available at <http://www.msnbc.msn.com/id/35103622/ns/world_news-haiti/>.

⁷⁰ *Farmer Testimony* (n. 68), July 27, 2010.

⁷¹ See, e.g., "Camp Management Terms of Reference," IASC CCCM Cluster (providing a list of key responsibilities for agencies taking charge of the management of displacement camps in emergency settings).

This is the standard approach for the humanitarian system in situations where IDPs have created spontaneous settlements; model “terms of reference” and guidance exist for such operations.⁷² CMAs are usually large INGOs with significant experience running IDP or refugee camps in emergencies around the world.⁷³ Despite the involvement of a wide range of major INGOs, there was an insufficient number of agencies available to manage the more than 800 camps in the city.⁷⁴

INGOs work on a project model, through which they endeavor to service a specific grouping of camps or households with a set package of goods and services, usually measured through project indicators. These indicators are, in turn, greatly influenced by industry standards as set out in Sphere. To ensure that they can both ensure minimum quality standards and measure the impact of their programming, INGOs tend to prefer to work in locations and with populations where they can meet their minimum standards for service. When asked to define the term “indicator” in relation to their work in Haiti, a number of respondents to the online survey replied by describing project indicators.⁷⁵ These results mirror the findings of Monika Krause, whose empirical research on humanitarian INGOs found that relief is organized according to the “logic of project management,” which:

draws a boundary around an agency’s responsibility. Quantifiable results for chosen beneficiaries are maximized at the expense of the overall affected population. Those who are hardest to help often receive no assistance at all.⁷⁶

This dynamic has been critiqued as “humanitarian containment” by Helen Young and her colleagues in a study examining assistance in Darfur, since it limits the positive impact that humanitarians can have on the entire affected population by concentrating it on specific populations within project areas.⁷⁷ There is evidence of this dynamic in Haiti, where well-serviced camps sit alongside unmanaged camps. Because CMAs are judged according to their ability to meet indicators of minimum quality and accountability in relation to a project’s beneficiaries, the incentives are to prioritize quality over coverage. Failures of quality and coverage *within* project sites are traceable to specific CMAs; failures of coverage for the entire disaster or affected population are traceable to the amorphous “humanitarian system” as a whole, where accountability is dispersed. Although the humanitarian system has

⁷² See, e.g., *ibid.*

⁷³ “Displacement Tracking Matrix Summary Report,” IASC CCCM Cluster Haiti, June 2010.

⁷⁴ See, e.g., “Haiti, Earthquake Situation Report #34,” OCHA, April 16, 2010, available at <<http://onerresponse.info/Disasters/Haiti/Coordination/publicdocuments/OCHAsitrep34.pdf>>, 2 (noting that less than a quarter of these camps had camp management agencies).

⁷⁵ See, e.g., Response No. F1.13 (“[I]t is a way of quantifiably measuring specific project outputs and outcomes”); see also Responses No. F1.32, F1.1, F1.6, F1.16, F1.25, F1.48, F1.57, and F1.62.

⁷⁶ Krause, “The Logic of Relief” (n. 55), 14, 156.

⁷⁷ See Helen Young, Abdul Monim Osman, Yacob Aklilu, Rebecca Dale, Babiker Badri, and Abdul Jabbar Abdullah Fuddle, “Darfur: Livelihoods Under Siege,” Feinstein International Famine Center, June 2005, available at <<http://idp-key-resources.org/documents/2005/d04507/000.pdf>, 117>.

developed methods for coordinating work across sectors in the past several years—notably through the use of the Cluster Approach, which has allowed for improved attention to coverage—the quality–coverage tension arises from a structural problem that cannot be solved by better coordination alone.⁷⁸

The impact on residents of unmanaged camps is significant: without agencies coordinating aid, goods and services are delivered irregularly, if at all.⁷⁹ A survey of 90 families in six different IDP camps in and around Port-au-Prince, conducted by a coalition of groups in February 2010, found “sporadic, infrequent and apparently arbitrary” food distributions, contaminated water sources, and radically inadequate shelter.⁸⁰ A randomized study of 1 in 8 camps carried out in summer 2010 concluded that “[d]ata shows [*sic*] that camps with NGO management are far better serviced than camps without management.”⁸¹ In addition, despite efforts to mitigate the risks, some camps were still precariously situated in then-dry riverbeds, flood plains, and areas at high risk of landslides as of July 2010.⁸² By June 2010, an estimated 66 percent of all IDP households in Port-au-Prince were in locations with CMAs.⁸³ This meant that a striking 34 percent of households were in sites where no agency was ensuring the delivery of coordinated services. While CMAs might be able to boast fulfillment of Sphere indicators within their zones of work, the humanitarian system was failing large swathes of the IDP community six months after the earthquake.

Impossible indicators?

Soon after the disaster response began, it became apparent to the experienced CMAs that Sphere indicators would be difficult and often impossible to meet in the spontaneously settled sites where they were working in Port-au-Prince. For example, at the time of the earthquake, Sphere’s shelter indicators specified that “[t]emporary planned or self-settled camps are based on a minimum surface area of 45m² for each person.”⁸⁴ The fact that all of the original IDP camps were

⁷⁸ A recent evaluation of the Cluster Approach credits the system for better sectoral coordination and improved attention to coverage and “gap-filling.” See Abby Stoddard, Adele Harmer, Katherine Haver, Dirk Salomons, and Victoria Wheeler, “Cluster Approach Evaluation,” November 2007, available at <<http://www.humanitarianoutcomes.org/pdf/ClusterApproachEvaluation.pdf>>, 7–8.

⁷⁹ See “Haiti: Six Months After,” UN OCHA, July 12, 2010, 7 (stating that in spontaneous settlements, “services are intermittent at best”).

⁸⁰ “Neglect in the Encampments: Haiti’s Second-Wave Humanitarian Disaster,” Lamp for Haiti, Earl Macke School of Law at Drexel University, Institute for Justice & Democracy in Haiti, Bureau des Avocats Internationaux, and Lawyers’ Earthquake Response Network, March 23, 2010, available at <<http://ijdh.org/archives/10671>>, 1.

⁸¹ Schuller, “Unstable Foundations,” (n. 10), 18 (finding dramatically better outcomes for managed camps in terms of water services, health care, and overall conditions).

⁸² See “Haiti: Six Months After,” UN OCHA, July 12, 2010, 7 (noting that some camps were at high risk of flooding and destruction from storms, flooding, and landslides).

⁸³ By June 2010, data revealed that 100 percent of households in the richest part of the city—Pétionville—lived in sites with camp management, while in the city’s largest slum, Cité Soleil, only 22 percent of households benefitted from camp management structures “Displacement Tracking Matrix Summary Report,” IASC CCCM Cluster Haiti.

⁸⁴ Sphere Project 2004 ed (n. 37).

spontaneously settled, the extent of the destruction in Port-au-Prince, the population density of the city before the earthquake, and the relative lack of open land, combined to make this area indicator an unrealistic goal within many of the city sites.⁸⁵ “Conditions in Port-au-Prince are difficult,” a Sphere consultant reported in relation to the shelter space indicator.⁸⁶ “When agencies say it’s so difficult to do Sphere here, I agree. It’s incredibly difficult to follow these standards in that environment.”⁸⁷ However, the consultant emphasized that this difficulty should not become an excuse to ignore Sphere, counseling humanitarians not to “throw it out completely.”⁸⁸

One option is for humanitarians to “contextualize” universal indicators when they are impossible to meet for reasons specific to the setting or circumstances of the disaster. This term is frequently used by humanitarians to mean that an indicator may need to be relaxed to fit the situation.⁸⁹ This option was brought up by numerous survey respondents: one respondent specifically noted that the Sphere surface area indicator needed to be “contextualized” for use in Haiti, while several others emphasized the need to contextualize indicators more generally in Haiti.⁹⁰ The 2004 edition of the Sphere *Handbook* addressed this issue as follows:

[T]here is inevitably a tension between the formulation of universal standards and the ability to apply them in practice. Every context is different. In some instances, local factors may make the realisation of all standards and indicators unattainable. When this is the case, the gap between the standards and indicators listed in the handbook and the ones reached in actual practice must be described, and the reasons for it and what needs to be changed must be explained.⁹¹

One of the major issues that experts raised in interviews when asked about contextualization was the need to adjust the Sphere indicators downward if using them in an unmodified form would mean that program beneficiaries would be better off than those in the general community.⁹² The 2004 Sphere *Handbook* addresses the risk of this welfare gap: “[I]t can be the case that the Minimum Standards exceed normal everyday living conditions. Since this can give rise to resentment, local conditions must be taken into account, and programs should always be designed with equality of the affected and surrounding populations in mind.”⁹³ In Haiti, for example, where many non-displaced Haitians cannot access

⁸⁵ A number of respondents to the online survey commented that the shelter space indicators, among others, were impossible to meet in Port-au-Prince. Response No. F3.18 and F3.24 (shelter space indicator) and F3.4, F3.9, F3.31, F3.22, F3.23, F3.30 (other indicators or unspecified indicators).

⁸⁶ “Haiti: Humanitarian Best Practice—Dignity, not Just Digits,” IRIN, April 9, 2010, available at <<http://www.irinnews.org/Report.aspx?ReportId=88752>>.

⁸⁷ Ibid.

⁸⁸ Ibid.

⁸⁹ Interviews Nos 3 and 4.

⁹⁰ Response No. F3.18. See also F3.24 (shelter indicator impossible to meet in Port-au-Prince), and F3.4, F3.9, F3.31, F3.22, F3.23, F3.30.

⁹¹ Sphere Project 2004 ed (n. 37), 8–9.

⁹² Interview No. 3.

⁹³ Sphere Project 2004 ed (n. 37), 14.

15 liters of water per person per day without assistance, contextualization might mean adjusting the water quantity goal downward within IDP camps. Otherwise, the better services offered by CMAs in camps could be a source of tension between IDPs and the general community.⁹⁴ Another option suggested by one expert is to provide services—here water—to the whole affected population.⁹⁵ This suggestion may have limited impact in situations where coverage of even those displaced by the earthquake is inadequate.⁹⁶

This contextualization process, seen by humanitarians as an important tool for adjusting to local conditions, could also be seen as a problematic departure from the putative rights-based nature of the minimum standards, especially since the 2004 *Handbook* makes reference to the “core content” of the right to water.⁹⁷ As one respondent to the online survey said, “I found in my agency that NO ONE followed these standards [including Sphere]! As if the catastrophic conditions in Haiti created a negligible window to apply any standards.”⁹⁸ Other respondents underscored the importance of the rights-based approach in the Haitian context: one opined that the rights-based approach was important following five hundred years of “rights-violation based approaches,”⁹⁹ while several others indicated that the failure of past international interventions in Haiti underlines the need for the rights-based approach.¹⁰⁰

“Fake” victims?

Because most emergency aid is targeted at those defined as “disaster-affected,” humanitarians working in very poor settings worry not only about conflict erupting between those receiving services and those who are ineligible, but also about creating perverse incentives relating to eligibility criteria. Reports have surfaced suggesting that, in Haiti, such incentives were resulting in “ghost camps” where “fake” victims set up shanties to obtain aid during the day but returned to their homes in the evening.¹⁰¹ Such “ghost camps” challenge the ability of humanitarians to implement programming according to two of the main pillars of good

⁹⁴ Interview No. 3. Dr. Paul Farmer recently reported that “by some reports, [water insecurity] has lessened since the earthquake,” *Farmer Testimony* (n. 68), July 27, 2010.

⁹⁵ Interview No. 3.

⁹⁶ The disaster-affected population reportedly received about 5 liters of water per day during the first three months of the response. See François Grünewald, Andrea Binder, and Yvio Georges, “Inter-Agency Real-Time Evaluation in Haiti: 3 Months after the Earthquake” August 31, 2010, 54 (hereinafter Grünewald et al., “Inter-Agency RTE”).

⁹⁷ Sphere Project 2004 ed. (n. 37), 55.

⁹⁸ Response No. F3.8. Another respondent added that “The emergency is so overwhelming that almost all (if not all) of the services delivered are substandard, so many people don’t really seem to have SPHERE in mind anymore. Or they do, but in a long term perspective.” Response No. F7.5.

⁹⁹ Response No. F8.1.

¹⁰⁰ Responses No. F8.3, F8.4, F8.8, and F8.13.

¹⁰¹ Jessica Leeder, “Fake Encampments Spring Up in Haiti,” *Globe and Mail* (Canada), May 8, 2010. Beverly Bell, “Citizen Mobilization for Housing in Haiti,” *Huffington Post*, July 30, 2010, <http://www.huffingtonpost.com/beverly-bell/citizen-mobilization-for_b_664893.html>.

practice in assistance: ensuring that the disaster-affected communities receive services, and targeting those most vulnerable within that community for priority assistance. Understanding who is and is not “disaster-affected” is at the heart of the humanitarian endeavor, and divides humanitarianism from development work. Where development aims to improve the lot of entire national populations over long periods, humanitarian assistance aims to improve the lot of specific populations affected by emergency events in the short term. Despite the importance of this distinction, the Inter-Agency Standing Group reflected on its problematic nature in Haiti at the six-month mark:

The ability of the humanitarian community to identify the most vulnerable people in need of assistance has been, and continues to be, a major challenge. The underlying poverty and vulnerability across Haiti renders the qualification of “directly-affected by the earthquake” somewhat irrelevant in any case, considering that almost everyone in the country has been affected in some way. The food security sector is an example of this: although 69 percent of households in large IDP camps suffer from food insecurity for example, approximately 52 percent of households across the entire country are food insecure . . .¹⁰²

Despite the acknowledgment that the distinction between “disaster-affected” and “not disaster-affected” is “somewhat irrelevant,” humanitarians still use this categorization to determine eligibility for limited quantities of aid. One study demonstrated that:

displaced populations in official camps had higher quality facilities and services than those staying with family. For example, those registered in camps have a better chance of living in a waterproof shelter, accessing a latrine or flush toilet, and seeing security patrolling the camps.¹⁰³

Further, vulnerability operates even in very poor disaster-affected communities, and evidence shows that humanitarians may have failed to provide adequate services to some of the marginalized populations in Port-au-Prince. For example, by June 2010, data revealed that 100 percent of households in one of the richest parts of the city—Pétionville—lived in sites serviced by CMAs, while in the city’s most well-known slum, Cité Soleil, only 22 percent of households benefitted from CMAs.¹⁰⁴ A survey conducted by a Haitian think tank in Cité Soleil in February 2010 found that less than 28 percent of the 962 respondents had received assistance by the time of the survey.¹⁰⁵ Monika Krause notes that “[p]opulations *not* served or needs *not* fulfilled are never analysed as part of the actually existing system of relief.”¹⁰⁶

¹⁰² “Response to the Humanitarian Crisis in Haiti Following the January 12, 2010 Earthquake,” IASC, July 2010, available at <<http://www.interaction.org/document/iasc-response-humanitarian-crisis-haiti-additional-resources>>.

¹⁰³ See Rencoret et al., *Haiti Earthquake Response* (n. 67), 31 (citing Louise Ivers, Amanda Schiff, Douglas Jones, James Evans, and Marc Zissman, Assessment and analysis of Haitian HADR data (April 2010)).

¹⁰⁴ “Displacement Tracking,” IASC CCCM Cluster Haiti, June 2010.

¹⁰⁵ “Voices from the Shanties,” Inured, March 2010 (n. 9), 2. See also Schuller, “Unstable Foundations” (n. 10), 19 (finding dramatic differences in services and conditions in camps in, for example, Cité Soleil and Pétionville).

¹⁰⁶ Krause, “The Logic of Relief” (n. 55), 40.

Humanitarians have devised systems to monitor—at least at a very general level—the urgent needs of the entire affected population, including those outside managed camps. However, this monitoring has been insufficient to close the gaps it reveals.

By July 2010, the humanitarian system concluded that although IDP numbers appeared to be fluctuating, since there had not been any major aftershocks or destructive flooding or landslides, “it seems reasonable to assume that additional IDPs arriving at settlement sites and requesting humanitarian assistance have not been directly-affected in terms of destruction of their homestead and are vulnerable for other reasons.” If this is the case, such “additional IDPs” need not be prioritized according to the traditional humanitarian criteria. Longtime Haiti activist Beverly Bell questions the need to divide the “real” victims from the “fake” in such a context:

What’s the standard for being a “real” victim? That one lost everything but the clothes on one’s back? That one is a corpse still lying, flattened, in one of many buildings across town that now serves as a mausoleum? And what would it mean if people’s daily lives were so devastated that they had to go to crowded, muddy, inhumane refugee camps for an upgrade?¹⁰⁷

According to Bell, the logic of emergency is what creates the “fake victim” dynamic, since it divides those “deserving” aid from those undeserving of assistance, and because it comprehends those in dire need of assistance—albeit for different reasons from those who lost homes in the earthquake—as perpetrating a fraud rather than maximizing their ability to access the goods and services they need and indeed have a human right to access. This logic also disguises a decision that has moral and ethical overtones—who should be eligible to obtain aid—with a set of calculations that appear technical, casting those attempting to avoid such line-drawing as fakers seeking to game the system.¹⁰⁸

Anthropologist Erica Caple James points to a different potential impact of such line-drawing: should humanitarians target only a certain category of Haitians considered “at risk,” such victims may “become subject to further stigma and resentment in their communities, as were victims of human rights abuses from the coup and postcoup years.”¹⁰⁹ Her work also suggests that “fake” victims be understood as appropriating a form of strategic agency. Writing about beneficiaries of aid projects in Haiti in the late 1990s, she explains that they “learned that benevolent aid is temporary, fleeting, and myopic” and responded by competing for the scarce resources available, sometimes using performance and commodification of their own victim status to access goods and services.¹¹⁰ But where is the humanitarian system left if the very concepts of “disaster-affected,” “victim,” and “vulnerability” begin to fall apart? Should humanitarians refocus their work to

¹⁰⁷ Bell, “Citizen Mobilization for Housing in Haiti” (n. 101).

¹⁰⁸ I am grateful to Greg Beckett for pointing this out.

¹⁰⁹ James, *Democratic Insecurities: Violence, Trauma, and Intervention in Haiti*, (n. 63) xxiii.

¹¹⁰ See *ibid.*, 223–69, 291.

target all poor Haitians, opening the beneficiary pool to many millions? Such a suggestion, though potentially promising, also carries hidden perils.

Helping beneficiaries, harming systems?

The perils of responding to the problem of extreme poverty by expanding humanitarian services lie in part in the tendency of humanitarians to create parallel systems when they act in “the uneasy boundary between humanitarianism and development.”¹¹¹ Humanitarians are concerned about the harmful impact such parallel systems can have on nascent or already-existing systems.¹¹² The 2004 *Sphere Handbook* includes the following statement about this issue in the “What is Sphere?” chapter: “In all contexts, disaster response should support and/or complement existing government services in terms of structure, design and long-term sustainability.”¹¹³ Sphere endeavors to address this concern in the health sector through indicators directly discouraging the creation or use of parallel systems.¹¹⁴ For example, the following indicators were included in relation to the health sector:

- Local health facilities are supported and strengthened by responding agencies.
- No alternate or parallel health facilities and services are established . . . unless local capacities are exceeded or the population does not have ready access to existing services. The lead health authority is consulted on this issue.¹¹⁵

Despite these promising indicators and efforts by humanitarians to address this issue by attending to what they call “early recovery,”¹¹⁶ one respondent to the online survey commented that the way humanitarians use Sphere tends to weaken systems thinking in favor of chasing targets:

Many of the Sphere standards, though supposedly qualitative in nature, are interpreted as qualitative numbers, and were not being achieved even before the earthquake. [T]here is such a focus on “achieving” these “targets” which are merely outputs and do not reflect impacts or qualitative changes that can actually be contradictory to outputs. [I] cite the massive influx of health NGOs and provision of free health care in the initial months that has undermined the existing health system, forced several hospitals to close down, is

¹¹¹ Peter Redfield, “The Verge of Crisis: Doctors without Borders in Uganda,” in Didier Fassin and Mariella Pandolfi (eds), *Contemporary States of Emergency: The Politics of Military and Humanitarian Interventions* (Brooklyn: Zone Books, 2010), 173, 174.

¹¹² For a discussion of how one humanitarian INGO negotiated the problem of contributing to the “aid economy” among other issues within the “uncertain zone of suffering” between crisis and normalcy, see *ibid.*, 173–4, 185.

¹¹³ Sphere Project 2004 ed. (n. 39), 7.

¹¹⁴ It is striking that the Health Services chapter has an entire section devoted to health systems and infrastructure. This is not true for other sectors such as water and sanitation or food and shelter. See *ibid.*

¹¹⁵ *Ibid.*, 261–2.

¹¹⁶ The Cluster Approach includes an Early Recovery Cluster. “Early Recovery Cluster Overview,” OneResponse, <<http://oneresponse.info/GlobalClusters/Early%20Recovery/Pages/default.aspx>>.

threatening the health system workforce and may do more harm than good in the long term.¹¹⁷

Indeed, a number of hospitals went bankrupt amid the free services.¹¹⁸ This is a much-noted phenomenon in Haiti, which existed even before the earthquake. Catherine Maternowska identified a pattern similar to that noted here in relation to family planning services in Haiti in the late 1980s and early 1990s, when US aid was funneled through NGOs as a way of avoiding the instability of the Haitian government.¹¹⁹ Paul Farmer has noted the same pattern in foreign assistance in relation to health, education, water, and housing in Haiti.¹²⁰ Similarly, writing about the international response to the hurricanes in Haiti in 2004, Gregory Beckett has noted that the international community responded to the emergency through services that could be understood as “palliative care,” leaving unchanged the government’s “continued lack of permanent capacity,” thereby integrating Haiti “in a highly dependent way, into a global system in which repeated interventions are authorized by the normalization of disaster and crisis.”¹²¹ An evaluation of the emergency response in Haiti found that the drive to assess using quantitative data “underestimate[s] the importance of analyzing existing capacities and constraints.”¹²²

By directly supplying services when the need is absolutely dire without implementing robust mechanisms for strengthening the public systems in Haiti, aid continues to de-capacitate Haitian systems, following the pattern discussed at the outset of this section. Recognizing the pattern, Haitians are alive to the fact that INGOs often control more money, employ the most well-qualified Haitians, and make more wide-ranging decisions that impact their lives on a daily basis than their own institutions. The result, Catherine Maternowska’s fieldwork suggests, in a country that is proud of its history of revolution against slavery, may be resistance to the very improvement projects introduced in the name of helping Haitians, which may often be seen as so many attempts to subordinate Haiti to international priorities and powers.¹²³

Participation, not politics

Recognizing the disempowering effects of disaster relief efforts, humanitarians have integrated indicators into Sphere to ensure beneficiary participation in programming. “Sphere is about people having rights to a life with dignity,” the Sphere

¹¹⁷ Response No. F11.1.

¹¹⁸ Grünewald et al., “Inter-Agency RTE” (n. 96), 50.

¹¹⁹ See M. Catherine Maternowska, *Reproducing Inequities: Poverty and the Politics of Population in Haiti* (New Brunswick, NJ and London: Rutgers University Press, 2006), 134–48.

¹²⁰ *Farmer Testimony* (n. 68), July 27, 2010.

¹²¹ Beckett, “The End of Haiti” (n. 61), 407.

¹²² Grünewald et al., “Inter-Agency RTE” (n. 96), 47.

¹²³ Maternowska identifies this pattern in relation to US-sponsored family planning. See Maternowska, *Reproducing Inequities* (n. 119), 148.

consultant quoted earlier said in relation to Haiti.¹²⁴ Sphere's common standards, where indicators concerning the process of providing relief can be found, include requirements for beneficiary participation and consultation.¹²⁵ Sphere is not meant to be (only) about chasing set targets; it is also about ensuring that those affected by the disaster have a say in the solutions to the most pressing problems.

Further, human rights advocates have emphasized the need to consult Haitians in spontaneous settlements, suggesting that such consultation is a human rights imperative and can assist INGOs in creating more effective systems of distribution.¹²⁶ Two humanitarian accountability experts working in Haiti agreed during interviews, noting that participation can improve efficiency and dignity at once.¹²⁷ However, an evaluation of the first three months of the humanitarian response in Haiti found that the international response to the earthquake was "highly exclusive," and that Haitians were "not consulted, informed or included in the design, planning, and implementation" of the relief effort.¹²⁸

One advocate complained that in Haiti, "the whole cluster [system] is ingeniously set up to insulate the people making the decisions from the information they need to make the decisions they are making."¹²⁹ Cluster meetings, through which humanitarian organizations coordinate their work on a weekly basis, were routinely held at the UN logistical base, far from the center of town and past a carefully guarded security gate known to turn away Haitians who arrived without specific invitations.¹³⁰

Still, 83 percent of those online survey respondents answering a question about the impact of indicators reported that using indicators and benchmarks "ensures the active participation of beneficiaries,"¹³¹ suggesting that process-related indicators are viewed as an effective tool. However, when humanitarian aid is delivered via projects that are not well connected to the larger, existing Haitian systems, the impact of the humanitarian commitment to "participation" can be severely limited. One of the main factors limiting its impact is the unit of analysis for participation: while INGOs include language about participation and accountability to disaster-affected communities as a whole, the operational reality is that agencies frequently set up participation and accountability structures tied to their projects or their agencies. Thus, the concepts of participation and accountability are limited by the logic of project management. As one advocate with decades of experience working in Haiti suggested, it would not be difficult to imagine methods for prioritizing Haitian participation in the *decision-making systems* of the humanitarian community, not only the project implementation structures. This would mean, for

¹²⁴ "Haiti: Humanitarian Best Practice," IRIN, April 9, 2010.

¹²⁵ Ibid.

¹²⁶ See "Neglect in the Encampments," Lamp for Haiti et al., March 23, 2010; *see also* Interviews Nos 16 and 19.

¹²⁷ Interviews Nos 18.1 and 18.2.

¹²⁸ Grünewald et al., "Inter-Agency RTE" (n. 96), 41.

¹²⁹ Interview No. 16.

¹³⁰ On-site visit by author to UN logbase (June 2010) (on file with author).

¹³¹ Online survey results (on file with author).

example, hiring Haitians to co-lead every Cluster and ensuring—as needed—that they are trained in the relevant skills and provided with an orientation to the humanitarian system; making sure that all Cluster meetings include Kreyòl translation; and reaching out to Haitian organizations to hear what their priorities are.¹³² Although some of these steps have been taken in Haiti, they have not been systematically pursued.¹³³

Moreover, evidence from the development field suggests that engaging populations in participation through projects can, among other effects, lead to manipulation of local knowledge for project ends; reinforce the power of those already powerful; be perceived as a method of exerting foreign control; and be structured in such a way that participation aimed at challenging existing power dynamics is made marginal.¹³⁴ Indeed, development analysts have suggested that “the language of empowerment masks a real concern for managerial effectiveness.”¹³⁵ This risk seems inherent in the humanitarian context, which is already replete with management talk and concern about marrying efficiency to effectiveness. A recent analysis of the humanitarian situation in Haiti noted that there was “a striking disconnect between the objectives and plans of the international community and the Haitians.”¹³⁶ An evaluation of the earthquake response similarly concluded that coordination between the humanitarian community and Haitian civil society and government actors was poor.¹³⁷

The humanitarian commitment to political neutrality can lead agencies to ignore power relations and how they play out in the relief context. HAP recently documented this phenomenon within certain camp committees in Haiti, which it labeled as “Unaccountability Mechanisms” in some camps, where committees were essentially run by gangs.¹³⁸ One study conducted in summer 2010 found that “NGO-managed camps had worse governance indicators than those without.”¹³⁹

By approaching participation as a technical exercise, humanitarians “render technical” the very political issue of how poor people in Haiti can take an active part in their country’s rebuilding.¹⁴⁰ One Haitian NGO recently observed that during this time of reconstruction:

¹³² Interview No. 16.

¹³³ Interviews Nos 16, 19, and 17.

¹³⁴ See Bill Cooke and Uma Kothari, “The Case for Participation as Tyranny,” in Bill Cooke and Uma Kothari (eds), *Participation: The New Tyranny?* (London: Zed Books, 2001), 1.

¹³⁵ *Ibid.*, 14.

¹³⁶ Lama Khouri-Padova, *U.N. Peacekeeping Best Practices Unit, Haiti: Lessons Learned* 8 (2004), available at <<http://smallwarsjournal.com/documents/haitilessonslearned.pdf>>.

¹³⁷ “Haiti: From the Ground Up,” Refugees International, March 2, 2010, available at <http://www.refugeesinternational.org/sites/default/files/030210_haiti_groundup.pdf>.

¹³⁸ “Camp Committees in Haiti: Un-Accountability Mechanisms?,” HAP, May 28, 2010, available at <<http://www.hapinternational.org/news/story.aspx?id=175>>.

¹³⁹ Schuller, “Unstable Foundations” (n. 10), 31.

¹⁴⁰ For a discussion of the concept of “rendering technical,” see Tania Murray Li, *The Will to Improve: Governmentality, Development, and the Practice of Politics* (Durham and London: Duke University Press, 2007), 7–9.

if public policy exists, it is defined by “experts” from international financial institutions (IFI), representatives of the diplomatic missions and international organizations accredited in Haiti, accompanied by some Haitian bureaucrats/technocrats and politicians and by so-called members of civil society.¹⁴¹

Is it wise, one might ask, for Haitians to spend a great deal of (unpaid) time consulting with INGOs about how best to implement a project, or would that time be better spent engaging in civic action and working to influence the country’s future?¹⁴² This is an especially complex question given that INGOs control so many of the reconstruction resources, making engagement with them crucial. Still, participation in specific assistance projects is not the same thing as direct engagement with the major institutions—such as the Commission Intérimaire pour la Reconstruction d’Haïti—that were setting priorities, disbursing funds, and coordinating reconstruction activities. Indeed, focus on participation in INGO programming may divert attention from the lack of participatory mechanisms built into the Commission and its counterpart bodies charged with governing the reconstruction in Haiti.¹⁴³

As one humanitarian respondent to the online survey with years of experience in Haiti remarked:

I was at first blindly optimistic that my work with my agency would provide an opportunity to enhance a rights-based perspective, especially in the area of human rights [violations] but this was not the case. The agencies’ mandate while rights based in theory was not fully functional in the emergency setting. Haiti is a country with a very strong democratic popular movement but no one in the international agencies dared to look at this as an asset.¹⁴⁴

No indicators for protection

While there are very real problems introduced through the audit culture of indicators, the problems are not always about an overemphasis on metrics. Advocates working in Haiti have also underscored the *reluctance* of the humanitarian system to produce certain kinds of data: numbers about civil and political human rights violations occurring in the IDP camps. In Haiti, the silence of Sphere and other indicators projects in relation to such protection issues has been perhaps most palpable in connection with gender-based violence and forced displacement. Before the 2011 edition, Sphere included only one indicator specifically addressing gender-based violence (GBV): health-service providers are instructed to “prevent and manage the consequences of gender-based violence.”¹⁴⁵ While gender-based violence has been

¹⁴¹ Roland Bélizaire, “Politique Publiques en Haïti: à Quand la Rupture avec la Dépendance?” [“Public Policy in Haiti: When Will We Break with Dependency?”], PAPDA, July 2010, 19 (translation by author).

¹⁴² See, e.g., Fabiola Córdova, “Haitians Need Their Own Voice,” *Americas Quarterly Blog*, July 28, 2010, available at <<http://www.americasquarterly.org/node/1741>>.

¹⁴³ I am grateful to Greg Becket for pointing this out.

¹⁴⁴ Response No. F8.8.

¹⁴⁵ Sphere Project 2004 ed. (n. 37), 288.

addressed in relation to numerous principles in the new 2011 protection chapter, there are no indicators in this chapter, meaning that Sphere does not contain standards concerning monitoring cases of rape or sexual violence themselves.¹⁴⁶ Instead, the chapter includes the following guidance note on “Monitoring and Reporting” with reference to the principle of “Protection from Violence and Coercion”:

Humanitarian agencies should consider their responsibility to monitor and report grave violations of rights. They should also consider advocating for the rights of affected populations with relevant authorities and actors by reminding them of their obligations. They may use different modes of action including diplomacy, lobbying and public advocacy.¹⁴⁷

As one advocate explained, humanitarians “are very reluctant to us[e] any numbers at all” in relation to rape.¹⁴⁸ This is striking in a context where data and metrics otherwise abound: “They quantify the number of tarps they give,” said the rights advocate, but don’t count the number of sexual attacks that take place within their camps.¹⁴⁹ While the humanitarian system emphasized that “one rape is too many,” a Haitian women’s rights organization that had been collecting its own statistics reported that it had documented 230 cases of rape in the first two months following the earthquake—in only 15 of the more than one thousand IDP camps.¹⁵⁰

Similar problems occur in relation to forced displacement. Private landowners have ousted IDPs from their land in the night, sometimes with the assistance of armed gangs, and in some cases helped by the Haitian National Police.¹⁵¹ One advocate working on this issue was mystified about what she characterized as the passive way that INGOs greeted forced displacement.¹⁵² Theorizing that perhaps the INGO workers were inured to suffering after witnessing so much disaster in country after country, she explained that a variety of INGOs were well aware of forced displacements being carried out by or at the behest of landowners.¹⁵³ In some cases, landowners were preventing INGOs from distributing food on their land.¹⁵⁴ There is no lack of legal guidance for humanitarians regarding the human rights of IDPs, including specific rights concerning alternatives to displacement, as

¹⁴⁶ See *ibid.*, 40, 56, 113, 170, 209, 269, 284–5, 289–90 (chapter introductions and guidance notes mentioning GBV).

¹⁴⁷ Sphere Project 2011 ed. (n. 38), 38.

¹⁴⁸ Interview No. 19.

¹⁴⁹ *Ibid.*

¹⁵⁰ See “Our Bodies Are Still Trembling,” Institute for Justice and Democracy (n. 12) in Haiti et al., 4, 20 (quoting the Gender-based Violence Sub-cluster and reporting that Haitian women’s NGO KOFIVIV had documented 230 rapes in 15 camps in the first two months following the earthquake). For a summary of the security context for women in the IDP camps, see, “Rapport sur la Sécurité dans les Camps,” July 18, 2010, KOFIVIV, available at <<http://ijdh.org/wordpress/wp-content/uploads/2010/08/KOFIVIV-Rapport-Sur-La-S%C3%A9curit%C3%A9-Dans-Les-Camps.pdf>> (explaining that IDPs have been forced to create informal security arrangements including the use of whistles as anti-rape mechanisms). See also Schuller, “Unstable Foundations” (n. 10), 22 (reporting on GBV in camps surveyed).

¹⁵¹ See, e.g., Mark Snyder, “Vanishing Camps at Gunpoint,” in Mark Schuller and Pablo Morales (eds), *Tectonic Shifts: Haiti Since the Earthquake* (Sterling: Kumarian Press, 2012), 143 (describing cases of violent forced evictions of IDPs); Speri, “Still Homeless from Haiti Earthquake” (n. 15).

¹⁵² Interview No. 17.

¹⁵³ *Ibid.*

¹⁵⁴ *Ibid.*

well as rights to security and dignity.¹⁵⁵ In addition to international legal materials, exhaustive operational guidance on how to ensure the protection of internally displaced persons¹⁵⁶ and persons affected by natural disasters, have been published.¹⁵⁷ Despite these materials, the project management logic appears to contribute to protection gaps: CMAs are responsible only for monitoring protection in camps under their management,¹⁵⁸ and the Protection Cluster is generally charged with monitoring the protection situation in Port-au-Prince.¹⁵⁹ The Protection Cluster has engaged in advocacy with the government of Haiti and managed to obtain a temporary moratorium on forced displacements.¹⁶⁰ However, reports indicate that evictions continued at the six-month mark; the Displacement Tracking Matrix reported “10 sites closed for eviction linked to private land” between May 3 and July 7, 2010.¹⁶¹ In a randomized study of one in eight camps in Port-au-Prince, Mark Schuller found that by summer 2010, 17 percent of camps had been closed through eviction.¹⁶² These bare data are emblematic of what is publicly reported concerning forced displacement by the humanitarian system. Human rights advocates expect there are many more cases than those which have been publicized or counted.¹⁶³ One advocate suggested that as a general matter, unmanaged camps have been “abandoned”¹⁶⁴ to their fate.

Humanitarian indicator projects may shift attention away from violations like these by failing to include relevant metrics for such abuses, and by emphasizing the importance of creating well-planned, indicator-compliant camps. The 2011 edition of Sphere includes guidance notes on freedom of movement in the protection chapter, including one that clearly states that “[p]eople should not be forced to stay in, or go to, a place that is not of their choice (such as a camp) nor should any other unreasonable restrictions be placed on their movement.”¹⁶⁵ Despite this clear statement, the lack of indicators may be in part to blame for the relative silence of the humanitarian community in the face of forced displacement.

¹⁵⁵ See, e.g., UN OCHA Guiding Principles on Internal Displacement, February 11, 1998, UN Doc. E/CN.4/1998/53/Add.2.

¹⁵⁶ “Handbook for the Protection of Internally Displaced Persons,” Global Protection Cluster Working Group, March 2010, available at <http://onerresponse.info/GlobalClusters/Protection/Documents/IDP%20Handbook_FINAL%20All%20document_NEW.pdf>, 137–45.

¹⁵⁷ “Human Rights and Natural Disasters: Operational Guidelines and Field Manual on Human Rights Protection in Situations of Natural Disaster,” Brookings-Bern Project on Internal Displacement, March 2008, available at <http://www.brookings.edu/~media/Files/rc/reports/2008/spring_natural_disasters/spring_natural_disasters.pdf>, 8.

¹⁵⁸ See “Camp Management,” IASC CCCM Cluster (n. 46).

¹⁵⁹ See “Terms of Reference,” Haiti Protection Cluster (N.D.) (specifying that the Cluster will, *inter alia*, “[i]dentify protection issues and gaps and use that information to: 1. advocate with the Government and support it through capacity building initiatives aimed at preventing and addressing protection concerns; 2. coordinate agencies to respond to identified gaps”).

¹⁶⁰ See Ansel Herz, “As ‘Temporary’ Camps Linger, Tensions Rise with Haitian Landowners,” *IPS News*, June 9, 2010, <<http://www.ipsnews.net/news.asp?idnews=51774>>.

¹⁶¹ “Displacement Tracking Matrix Summary Report,” IASC CCCM Cluster Haiti (n. 73).

¹⁶² Schuller, “Unstable Foundations” (n. 10), at Exec. Sum; see also *ibid.* at 23–5 (exploring the evictions issue in greater detail).

¹⁶³ Interview No. 17.

¹⁶⁴ *Ibid.* ¹⁶⁵ Sphere Project 2011 ed. (n. 38), 39.

Although the two dynamics may not be directly linked, this relative silence, coupled with the desire to move IDPs into planned camps, may create a situation in which forced displacement is inadvertently tolerated in the hopes that moving IDPs to more Sphere-compliant camps will ensure their longer term safety despite the unlawful nature of the movement and attendant human rights violations. The stated objective of not encouraging IDPs who fled to the countryside to return to Port-au-Prince could also be part of this dynamic.¹⁶⁶

As of July 2010, only one major planned, reportedly Sphere-compliant camp had been constructed near Port-au-Prince: Corail-Cesselesse, where “a few hundred acres were picked out for the city’s first, long-awaited relocation camp. The UN and US military construction teams flattened and graded the land for a camp of deluxe ‘ShelterBox’ tents. About 5,000 residents of the Pétienville Club golf-course camp, run by actor Sean Penn, were bused in.”¹⁶⁷ This camp was said to offer food, potable water, sanitation, health care, education, child protection, cash-for-work, and disability services.¹⁶⁸ Tents were laid out in grids, and residents reportedly were given a hygiene kit and two weeks of food rations upon arrival, as well as \$50 cash grants two weeks later.¹⁶⁹ However, months after the camp opened, residents reported that the camp—while neat and orderly in many respects—was like a desert baking in the sun without any shade.¹⁷⁰ In addition, while assistance may have been plentiful at first, many had abandoned Camp Corail by July 2010 because services were inadequate and it was so far away from the few livelihood opportunities and schooling options available in the city.¹⁷¹

Fears that (more) Sphere-compliant camps might draw IDPs or create tensions with the surrounding community appear to have proven true, though with a twist: “Squatter” camps have been set up alongside Corail-Cesselesse, and wealthy land-owners are suspected of being behind intimidation attacks during summer 2010 on these unmanaged camps.¹⁷² Including protection indicators in the Sphere *Handbook*, which would require humanitarians to measure such dynamics, might be a significant step forward in this setting.¹⁷³

¹⁶⁶ This suggestion was made by Melinda Miles in a field report to the Haiti Response Coalition, which has been reporting on conditions in the unmanaged camps. See Melinda Miles, “Haiti’s Answer for Six Months and Sixty Years,” *Haiti Response Coalition*, August 12, 2010, <<http://www.lehaitilive.org/news-english/2010/8/12/haitis-answer-for-six-months-and-sixtyyears.html>> (“The aid community has intentionally left the inhabitants of camps without access to better basic services. It is a strategy underway right now to avoid luring people back from the countryside with the promise of services in the camps.”).

¹⁶⁷ “Disputes Over Land Stall Haiti’s Recovery,” *Associated Press*, July 11, 2010, <<http://newsone.com/world/associatedpress2/disputes-over-land-stall-haitis-recovery/>>.

¹⁶⁸ “CCCM Update 3,” IASC CCCM Cluster, May 7, 2010.

¹⁶⁹ Ibid.

¹⁷⁰ See “Haiti Cherie,” TransAfrica Forum, July 2010, available at <http://www.transafricaforum.org/files/Haiti_Report.pdf>, 7–8.

¹⁷¹ Ibid.

¹⁷² “Disputes Over Land Stall Haiti’s Recovery,” *Associated Press* (n. 167). Further, news accounts indicate that landowners are seeking significant sums for the land that IDPs have begun to claim. Ibid.

¹⁷³ For a discussion of the new protection chapter, to be included in the 2011 version of the Sphere *Handbook*, see Satterthwaite, “Rights Based Humanitarian Indicators,” (expanded version) *N.Y. U. Journal of International Law & Politics* 43 (August 2011): 865.

Conclusion

Rights-based standards and indicators have become central to international emergency action. Integrating specific understandings of human rights, these self-regulation projects attempt to codify the “lessons learned” by the humanitarian community following a series of humanitarian failures, especially those in Somalia, Bosnia, and Rwanda. Using the language of management and the logics of audit, the indicator projects encapsulate the knowledge and expertise of professional humanitarians across key sectors. As such, they represent an impressive collaborative effort by a transnational network to learn lessons and evolve professional practices in an incredibly challenging environment. At the same time, the standards and indicators tend to “render technical” the core tensions that led to their creation, displacing rather than solving those issues. Indeed, current debates over what to measure and monitor via these indicators may be understood as symptoms of the unresolved tensions inherent in the humanitarian endeavor today.

These unresolved tensions lurk below the surface of the humanitarian response in Haiti. For example, indicator-related issues concerning coverage and scope of services, the management of the dividing line between emergency relief and development work, and the role of the operational community in ensuring protection of disaster-affected communities, have been discussed as technical problems in Haiti. Meaningful discussion of the issues at the heart of post-disaster Haiti is thus sidelined.

In the end, data and measurement are not the problem. Indeed, more precise and accurate data is essential to ensuring that assistance to disaster-affected communities is carried out in ways that are more effective, rights-enhancing, and accountable. The problem lies in what is effaced once a data set becomes fixed as an “indicator”: once codified as such, the debates underlying what it means to provide quality assistance, to be accountable, and to ensure human rights tend to retreat from view. Indicators appear measurable, precise, and orderly. As such, they seem to be the inevitable product of a scientific method, instead of one among many possible agreements at the end of a human process. More explicit attention to the tensions elided by these technical discussions could allow humanitarians—working together with affected communities themselves—to continue to harness the power of numbers and audit while also resisting the means through which indicators tend to disproportionately empower those with the greatest technical expertise—those governing at a distance.

Impact Investment Indicators: A Critical Assessment

*Sarah Dadush**

Commercial investing and philanthropy have traditionally been separate enterprises, with different regulations, policies, and moral justifications pertaining to each. Increasingly, however, with the emergence of innovative forms of investment such as “impact investing,” the re-labeling of donors as “social investors,” and the creation of hybrid corporate forms that adopt “double bottom line” business models, the distinction between commerce and philanthropy is being simultaneously blurred and bridged.

The incorporation of social issues into commercial investing and of market values into philanthropy is transforming how solutions to social and environmental problems are being developed and financed globally. This chapter seeks to show that indicators for assessing social and environmental impacts play a vital role in laying the foundations for socially oriented investing, on the one hand, and for market-oriented philanthropy, on the other. Further, indicators can contribute to changing the value structures that support both commercial investment and philanthropy. While this financing and value convergence may generate effective solutions for addressing social and environmental problems, it also risks narrowing the range of conceptual and practical tools available for solving those problems.

Impact investing serves here as a focal lens through which to examine the transformations outlined above. Impact investing is broadly defined as investing with the intent to generate social and environmental benefits, in addition to financial returns. In the next five to seven years, impact investments are expected to grow to over US\$500 billion.¹ The impact assessment framework composed of the Impact

* The author warmly thanks Eyal Benvenisti, Nehal Bhuta, Lorenzo Casini, Kevin E. Davis, Sapana Doshi, Angelina Fisher, Benedict Kingsbury, Felise Nguyen, and Richard Stewart for their insightful comments and support.

¹ See Monitor Institute, “Investing for Social and Environmental Impact: A Design for Catalyzing an Emerging Industry,” 2009, 9 (hereinafter Monitor Report). See B Lab, “GIIRS Q1 Progress Report,” 2011, available at <<http://www.giirs.org/about-giirs/progress-report>> (hereinafter Q1 Progress Report); see also Nick O’Donahoe, Christina Leijonhufvud, Yasemin Saltuk, Antony Bugg-Levine, and Margot Brandenburg, “Impact Investments, An Emerging Asset Class,” JP Morgan Global Research (hereinafter Emerging Asset Class), November 29, 2010.

Reporting and Investment Standards (IRIS) and the Global Impact Investing Rating System (GIIRS) provides the empirical focus of this chapter. IRIS establishes a freely accessible impact investing taxonomy or language for adoption by industry participants and beyond. Its ambition is to become the *de facto* reporting standard for impact investing, much like the International Financial Reporting Standards, which are used by thousands of businesses across the globe to report on corporate financial positions, and which have been officially adopted in over 120 jurisdictions.² IRIS provides definitions of impact investing terminology, as well as a menu of hundreds of metrics that investors and investees can use to track and report on social and environmental performance.

GIIRS adds a judgment function to the impact assessment process by aggregating IRIS-compliant data into numerical and star-based scores and ratings. Fee-paying companies and investment funds operating across the globe can obtain a GIIRS rating, and subscribers to the GIIRS service can access detailed industry benchmarks and sectoral analyses. Because it has a clear judgment and ordinal function, GIIRS more closely fits the definition of “indicator,” as laid out in the introductory chapter of this book, than does IRIS.³ However, because GIIRS draws so directly on the standards and metrics supplied by IRIS, the two are considered together as distinct but complementary tools for telling the same impact stories.

Both initiatives are US based and draw on financial and technical support provided by US entities, such as the Rockefeller and the Bill and Melinda Gates foundations, and a mix of non-profit (B Lab and Acumen Fund), business (e.g., Deloitte, Prudential Financial), and governmental (US Agency for International Development) organizations. Initial versions of IRIS and GIIRS were launched in 2010 and 2011, respectively. The development of these impact assessment tools involves iterative and consensus-driven processes that are designed to incorporate public feedback into updated versions. Thus, this piece examines indicators that are in the process of becoming established, rather than those already established.

This chapter identifies three key functions that indicators can play in governance; first, as interventions that influence individual decision-making; second, as interventions that influence governmental decision-making; and third, as interventions that influence the very process of decision-making and knowledge production. Each section explores a different governance function and concludes with an analysis of possible challenges to that function. Focusing on GIIRS and IRIS, we first examine how indicators, as collectors and producers of “market intelligence” can influence decision-making by socially or “mission-driven” investors and

² See Deloitte, “Use of IFRS by Jurisdiction,” available at <<http://www.iasplus.com/country/useias.htm>>.

³ As set out in Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, “Introduction: Global Governance by Indicators,” this volume, pp. 3–28, esp. 6: “An indicator is a named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis, such as countries or institutions or corporations), synchronically or over time, and to evaluate their performance by reference to one or more standards.”

investees. We then explore their influence on governmental decision-making, particularly where policy and regulatory frameworks are not yet established to govern a new area of activity, which is the case for impact investing. Finally, we consider how indicators can influence the very process of decision-making and knowledge production across two different problem-solving arenas, namely, commerce and philanthropy. In this last section, we describe impact indicators as “norm-vessels” that travel between commerce and philanthropy, transforming both in the process, and giving rise to questions about the importance of maintaining diversity within our problem-solving toolkits.

Indicators and market-building

Impact investing is currently establishing itself as a new industry and “asset class”;⁴ in this process, indicators play a crucial role as market-building devices that both inform and influence the decisions of investors and investees. Indicators communicate information about the state of the industry as a whole; they set benchmarks against which the performance of investees can be measured (internally or externally); they contribute to the establishment of industry standards and definitions; and they foster investor confidence and attract additional capital by enhancing transparency, comparability, and predictability. This section explains impact investing and the process of turning this growing area of activity into a new asset class. It describes two indicator-based systems—IRIS and GIIRS—that have been designed specifically to support impact investing. It concludes by highlighting some challenges faced by these systems.

What is impact investing?

Impact investments are defined as “investments intended to create positive impact beyond financial return.”⁵ Impact investing introduces “a new type of capital, merging the motivations of traditional investments and donations”⁶ for investors interested in using “for-profit investment to address social and environmental problems.”⁷ As such, impact investing is a primary example of the incorporation of social and environmental issues into commercial investing, and of market values into mission-driven funding, like philanthropy or international development assistance—where the primary purpose is to advance a socially beneficial mission.

⁴ See Emerging Asset Class (n. 1), 24–30.

⁵ Emerging Asset Class (n. 1), 7.

⁶ Emerging Asset Class (n. 1), 13.

⁷ Global Impact Investing Network (GIIN) Press Release, “New Industry Group Launched to Facilitate For-Profit Investing that Addresses Social and Environmental Challenges,” September 25, 2009, available at <<http://www.prnewswire.com/news-releases/new-industry-group-launched-to-facilitate-for-profit-investing-that-addresses-social-and-environmental-challenges-61447207.html>> (GIIN Press Release).

Proponents explain that impact investors and investees are driven by dual motives: a desire to achieve financial returns on the one hand, and a desire to achieve a social and/or environmental impact, on the other. While the systems for generating and assessing financial returns are well-developed and familiar to most businesses, systems for generating and assessing social and environmental returns are less so. At the outset, then, impact investing faces the challenge of identifying and quantifying social and environmental impact or “value.” In meeting this challenge, impact investing draws on Jed Emerson’s conceptualization of “the blended value proposition,” which states “that all organizations, whether for-profit or not, create value that consists of economic, social and environmental value components—and that investors (whether market-rate, charitable or some mix of the two) simultaneously generate all three forms of value through providing capital to organizations. The outcome of all this activity is value creation and that value is itself non-divisible and, therefore, a blend of these three elements.”⁸

One way to understand impact investing is as the next level of corporate social responsibility (CSR), or as CSR-*plus*: the onus on corporate entities is not only to avoid producing or contributing to social and environmental *harms*, which is the domain of conventional CSR, but also to produce social and environmental *benefits*. Conventional CSR uses negative screening to avoid investments in companies that produce social or environmental harms (e.g., tobacco, weapons, alcohol, gambling).⁹ In contrast, impact investing seeks actively to benefit society by funding companies that promise to generate social and/or environmental returns, as well as financial returns. It does this by expanding access to services or products for people in need (e.g., housing, health, and financial services) and/or through the adoption of climate-friendly and inclusive production processes (e.g., enhancing plant energy efficiency and purchasing green inputs from local providers).¹⁰

A central theme in impact investing is the focus on unlocking and channeling capital toward problem-solving initiatives that can address the world’s most pressing and intractable problems, such as poverty and climate change.¹¹ More specifically, proponents begin from the premise that the world’s “social balance sheet” is in deficit by many trillions of dollars, and argue that this amount is far too great for government assistance and philanthropy alone to supply.¹² Impact investing, itself largely driven by philanthropic entities such as the Rockefeller Foundation, aims to

⁸ Jed Emerson, “Blended Value,” available at <<http://www.blendedvalue.org>>.

⁹ Matt Krogh, Sustainable Industries, “Channeling Investment for Impact: New Rating System Helps Investors Move Beyond Responsibility,” April 29, 2010, available at <<http://blog.bcorporation.net/2010/04/channeling-investment-for-impact-new-rating-system-helps-investors-move-beyond-responsibility>>.

¹⁰ Emerging Asset Class (n. 1), 8.

¹¹ Antony Bugg-Levine and John Goldstein, “Impact Investing: Harnessing Capital Markets to Solve Problems at Scale,” *Community Development Investment Review* (Federal Reserve Bank of San Francisco) 5:2 (2009): 30 (hereinafter *Harnessing Capital Markets*).

¹² See Antony Bugg-Levine, “Impact Investing: A Global Perspective on an Emerging Capital Market, talk given at MaRS,” May 10, 2010, viewable at <<http://vimeo.com/10779881>> (hereinafter *MaRS*).

fill this funding gap.¹³ Those supporting the growth of impact investing underscore the importance of complementing philanthropic and public moneys with commercially sustainable and return-seeking funds: “unlocking institutional assets from the private banks, pension funds and insurance companies is where the big victory is going to be.”¹⁴

The following are examples of impact investing initiatives. In the energy sector, the (non-profit) fund E + Co invests seed and growth capital in renewable energy businesses in Africa, Asia, and Latin America: “E + Co’s business development support and investment capital serve to create energy businesses that mitigate climate change and reduce poverty while generating financial returns.”¹⁵ Accordingly, E + Co’s investment criteria are environmental and social, as well as financial.¹⁶ In agriculture, Root Capital (another not-for-profit, which receives significant support from Starbucks) lends to small rural businesses that would normally be viewed as being too small and risky to be investable and that produce items such as coffee, nuts, timber, and artisanal goods.¹⁷ The organization secures its loans by arranging fixed-price forward contracts from buyers.¹⁸ An increasing number of funds also provide debt or equity financing to social enterprises, which use business models to achieve positive and sustained social impacts, alongside financial returns, and target marginalized populations through increased employment or access to services, for instance.¹⁹

Capital markets are also tapped to finance projects that generate social benefits alongside positive financial outcomes in the global health setting. For example, the International Finance Facility for Immunization, launched in 2006, issues interest-bearing bonds in the capital markets for purchase by a range of institutional and retail investors, with the intention of raising \$4 billion by 2015 to fund immunization programs in developing countries. (So far it has raised about \$2.6 billion.) The bonds are secured by legally binding long-term commitments from donor governments.²⁰ As another example, Britain recently issued a Social Impact Bond, where returns on investment are linked to social performance metrics, such as a reduction in prisoner recidivism. Here the government commits to set aside a portion of the

¹³ In MaRS, Antony Bugg-Levine, head of the Impact Investing Initiative at the Rockefeller Foundation, referred to a deficit in the social balance sheet of trillions of dollars in capital needs, with only hundreds of billions at best coming from private charity and government.

¹⁴ Antony Bugg-Levine, Interview, @lliance for Philanthropy and Social Investment Worldwide, May 1, 2010, available at <<http://www.alliancemagazine.org/node/3283>>; see also the transcript of a debate between Matthew Bishop, author of *Philanthrocapitalism: How the Rich Can Save the World*, and Michael Edwards, author of “Just Another Emperor? The Myths and Realities of Philanthrocapitalism,” held April 9, 2009 (hereinafter Savior or Emperor), 7.

¹⁵ E + Co, “Approach,” available at <<http://eandco.net/about>>.

¹⁶ Ibid.

¹⁷ David Bornstein, “Filling the Gap Between Farm and Fair Trade,” *New York Times*, October 25, 2010.

¹⁸ Ibid.

¹⁹ For a definition of social entrepreneurs, see Alex Nicholls, “Social Enterprise and Social Entrepreneurs,” in Michael Edwards (ed.), *Civil Society* (2nd edn, Polity, 2009), 148. See also Report of the UN Commission on the Private Sector and Development to the Secretary General of the United Nations (2004) on Unleashing Entrepreneurship: Making Business Work for the Poor.

²⁰ International Finance Facility for Immunisation, available at <<http://www.iff-immunisation.org>>.

financial savings generated as a result of improved social outcomes for investor compensation.²¹

Impact investing draws on the legacies of microfinance and community development finance, both of which are characterized by social solution-driven models that can, and often do, also produce commercial returns. Building on the techniques developed in these areas, impact investments can take many forms, including equity, debt, working capital, lines of credit, and loan guarantees. Investments also target a broad spectrum of areas, including agriculture, clean energy, health, education, water, and infrastructure. Just as these investment areas are diverse, so are the types of investors involved. They include private foundations (e.g., the Rockefeller Foundation and the Omidyar Network), large-scale financial institutions (e.g., JP Morgan and Prudential) and commercial banks, private wealth managers, retirement fund managers, boutique investment funds, companies (e.g., General Mills and Starbucks), and development finance institutions (e.g., the Overseas Private Investment Corporation).²²

As should be clear from this list, the motivations for participating in impact investing are not the same for all of the investors involved: while some will prioritize social and environmental impacts over financial returns, others will place financial gains first;²³ while some will seek to engage with emerging markets in order to expand demand for their products among customers at the Base of the Pyramid (BoP), others will identify the achievement of international development goals as their main investment objective. Indeed, “[i]n some investors’ eyes, the coupling of the intent to create positive social impact with the pursuit of financial return is reason to expect lower returns from impact investments than from traditional investments. Others believe that financial return need not be sacrificed when social impact is being delivered and, due to the large under-penetrated market at the BoP, many impact investments should outperform traditional investments.”²⁴ Thus, impact investing attracts investors with a range of motivations, and industry-builders seem to be relatively agnostic as to the type of investor that brings resources to the table, so long as investments are channeled into businesses that are “designed with the intent to make a positive social and or environmental impact,” which “should be explicitly specified in company documents.”²⁵

²¹ Emerging Asset Class (n. 1), 20.

²² Emerging Asset Class (n. 1), 16.

²³ See Allen L. Hammond, William J. Kramer, Robert S. Katz, Julia T. Tran, and Courtland Walker, “The Next 4 Billion: Market Size and Business Strategy at the Base of the Pyramid,” World Resources Institute, 2007, 3: “Four billion low-income people, a majority of the world’s population, constitute the base of the economic pyramid [. . .] Their incomes in current US dollars are less than \$3.35 a day in Brazil, \$2.11 in China, \$1.89 in Ghana, and \$1.56 in India. Yet together they have substantial purchasing power: the BOP constitutes a \$5 trillion global consumer market.”

²⁴ Emerging Asset Class (n. 1), 30; see also Jed Emerson, “Steady Returns with Social Impact,” *Forbes.com*, September 29, 2009 and “Financial Innovation and the Poor: A Place in Society,” *The Economist*, September 25, 2009 (hereinafter *A Place In Society*) and *Harnessing Capital Markets* (n. 11), 39.

²⁵ Emerging Asset Class (n. 1), 14.

From threads to tapestry: Market weaving

Given that impact investing is made up of “different threads, born from different contexts,”²⁶ (microfinance, community development, health, education, etc.) it is not a new idea per se. However, weaving these threads into a “tapestry”²⁷ by linking these different activities together and establishing a market for them *is* new.²⁸ For a start, the potential dimensions of impact investing are becoming increasingly discernible and measurable: estimates suggest that in the next 5 to 10 years, the market for impact investments could represent 1 percent of all global assets under management (as calculated in 2008), over \$500 billion.²⁹

Promoters of impact investing are now better able to communicate certain key features about the industry, such as the average size of investment transfers, the size of the demand for impact investment opportunities, and the profile of impact investees, which was not the case until recently. A 2010 report by J.P. Morgan and the Rockefeller Foundation entitled “Impact Investing: An Emerging Asset Class” indicates that average impact investments will, at least initially, be quite small as compared with conventional commercial investing, in the region of \$1 million, with only very few deals hitting the \$10 million mark.³⁰ Growth in deal size is anticipated, however: “As impact investing matures and more institutional-scale investors with higher returns requirements enter the marketplace, we anticipate a proliferation of new investment funds being created, aggregating capital and increasing the size of investments that can be made. Average deal size will grow as the industry matures and fund vehicles facilitate larger deals.”³¹ From this description, it appears that growth in deal size is viewed as an important sign of progress for the impact industry itself, which raises the question: “Will the performance of this new asset class be judged on the basis of how much capital it attracts, how much is disbursed as impact investing, or how much it generates by way of social and environmental benefits—which may or may not be related to capital supply? If a combination, in what proportions?”

As concerns the recipients of impact investments—and their potential beneficiaries—the report also explains that while the hope is to eventually deploy impact investments to alleviate poverty amongst the very lowest-income households in developing countries, structural constraints dictate that the target demographic be made up not of the poorest of the poor, but of those making between \$500 and \$3,000 per year: “Constraints such as severely irregular cash-flows and the cost of distribution to these typically more remote populations will limit the ability of a profit-making business to deliver solutions to this sub-population.”³² Those at the very bottom of the pyramid would be served by governmental assistance and

²⁶ *Harnessing Capital Markets* (n. 11), 33.

²⁷ *Ibid.*

²⁸ *See A Place In Society* (n. 24).

²⁹ *See Monitor Report* (n. 1), 9.

³⁰ *Emerging Asset Class* (n. 1), 35. *See also* Paul Sullivan, “With Impact Investing, a Focus on More Than Returns,” *New York Times*, April 23, 2010.

³¹ *Emerging Asset Class* (n. 1), 35.

³² *Emerging Asset Class* (n. 1), 45.

philanthropy: "Impact investment can complement government and philanthropy by providing services to poor communities, thereby allowing government and philanthropy to concentrate their limited resources on reaching the poorest of the poor who cannot participate in market-based solutions."³³

Given that total giving to charitable organizations in 2010 amounted to approximately \$290 billion (with individuals contributing \$211 billion, foundations \$41 billion, and corporations \$15 billion), there is little doubt that impact investing flows could represent an important supplement to philanthropy.³⁴ In the face of concerns about the displacement of philanthropy—where no financial return is expected—by impact investments, advocates emphasize that the goal of impact investing is "not to cannibalize philanthropy" but rather to complement it:³⁵ "Impact investing can be a powerful complement to philanthropy and government efforts to address many issues, including relief for suffering communities, eradication of disease, stabilization of climate change, provision of basic social services such as housing and sanitation, and development of a green revolution in Africa."³⁶

Thus, within the impact investing landscape, philanthropic and public funds are reserved for those who are too poor to participate in market-based solutions, while impact investments kick in as beneficiaries move up the income ladder. Philanthropy may even serve a signaling function in this context, attracting attention to pressing problems that could potentially be addressed through market solutions.³⁷ The social and environmental problems being addressed by both philanthropic and market-based solutions are, however, largely similar. Should impact investing grow as anticipated, it could generate new fragmentations in the provision of public goods, such as health and housing, whereby philanthropic and governmental, as funding with no expectation of financial returns, provides for the poorest of the poor, while impact investors, who are at least partially driven by the prospect of financial gain, provide for those earning above a certain threshold. Such fragmentation could have important consequences for the delivery and distribution of public goods and the delegation of public functions, themes we will return to below.

One of the market features that remains unclear relates to the geographic allocation of impact capital. Specifically, what proportion of impact investments is expected to flow to emerging markets? Relatedly, given the social qualities of impact investing, could we imagine seeing these flows being included in calculations of international development assistance? How? While the J.P. Morgan and Rockefeller Foundation report mentions that impact investments will be split

³³ Emerging Asset Class (n. 1), 41.

³⁴ Charity Navigator, "Giving Statistics," available at <<http://www.charitynavigator.org/index.cfm?bay=content.view&cpid=42>>.

³⁵ See MaRS (n. 12); see also, Hope Neighbor, "Money for Good: \$120 billion Impact Investing Market Opportunity," *Tactical Philanthropy*, July 6, 2010, available at <<http://www.tacticalphilanthropy.com/2010/07/money-for-good-120-billion-impact-investing-market-opportunity>>.

³⁶ GIIN Press Release (n. 7).

³⁷ Antony Bugg-Levine, Interview, @alliance for Philanthropy and Social Investment Worldwide, May 1, 2010, available at <<http://www.alliancemagazine.org/node/3283>>.

between developed and emerging markets, and while the data and case studies contained in the report largely pertain to developing countries, there is little elaboration as to how the split will be determined.³⁸

Given that the estimated magnitude of flows would represent at least four times the current levels of official development assistance,³⁹ impact investing could have significant implications for the volume, composition, and allocation of capital flowing from rich to poorer countries, with attending implications for development.⁴⁰ Such implications relate to aid effectiveness, donor (/investor) coordination, as well as global policymaking.⁴¹ For instance, do impact investors operating in emerging economies espouse the achievement of the Millennium Development Goals? Do they align themselves with other development objectives? If so, which ones? And with what level of consultation and coordination? It is important to have clearer communication concerning the industry's aspirations for the geographic and development impact of impact investing capital.

Additional insight into the importance of emerging economies for impact investors would also help to better identify the intended beneficiaries of impact investments, to situate impact investing within the international development architecture for governance purposes, and to assess the performance of impact investing as a whole.⁴² This points to another feature of impact investing that is not entirely fleshed out: are there target beneficiaries for impact investments? We have a sense of who the investors are—though this is a wide range, given that both their motives and capital commitment levels differ considerably—and some sense of who the investees are—primarily mission-driven small and medium enterprises (SMEs) or for-profit entities whose primary purpose is to provide a social good—but there is little discussion as to who are the intended beneficiaries of impact

³⁸ Emerging Asset Class (n. 1), 21.

³⁹ Estimated at \$121 billion in Hudson Institute, "Index of Global Philanthropy and Remittances 2010," available at <http://www.hudson.org/files/documents/2010_IGP_Exec_Summary.pdf>.

⁴⁰ See two recent developments: OPIC Press Release, October 27, 2011: "In Historic Commitment to Impact Investing, OPIC Board Approves \$285 Million for Six Funds Catalyzing \$875 Million in Investments," available at <<http://www.opic.gov/news/pressreleases/2009/pr102711>> and GIIN Press Release, September 28, 2011: "USAID and Impact Investors Capitalize New Equity Fund for East African Agribusiness: A First of Its Kind Effort to Invest \$25 Million of Commercial Capital with a Private Investment Manager, Focused on Growing SMEs in the Agricultural Sector," available at <<http://www.rockefellerfoundation.org/news/press-releases/usa-id-impact-investors-capitalize-new>>.

⁴¹ See Jean-Michel Severino and Olivier Ray, "The End of ODA: Death and Rebirth of Public Policy," Center for Global Development, Working Paper No. 167, 2009, available at <<http://www.cgdev.org/content/publications/detail/1421419/>>.

⁴² See also Media Advisory: "Impact Investing Partnership with USAID, Rockefeller Foundation, Deloitte, and Prudential Financial to Support Entrepreneurs in the Developing World," *CSR Newswire*, April 26, 2010, available at <http://www.csrwire.com/press_releases/29468-Impact-Investing-Partnership-with-USAID-Rockefeller-Foundation-Deloitte-and-Prudential-Financial-to-Support-Entrepreneurs-in-the-Developing-World>, re USAID's participation in a \$6.5 million grant to support the development and use of GIIRS for emerging economies (Media Advisory); see also Sophia Grene, "New Social Impact Tool to Aid Investing," *The Financial Times*, May 2, 2010; see also *A Place in Society* (n. 24).

investments. Are the investees the intended beneficiaries, or is it impact receiving the goods and services provided via the investees? Could the vagueness concerning the identity of beneficiaries be inevitable given the breadth of activity areas covered by impact investing (agriculture, health, energy, housing, etc.)? What are the consequences of this vagueness for effectively assessing the impact of impact investing?

Another illustration of the evolution from individual threads to tapestry is the establishment of a trade association for impact investors called the Global Impact Investing Network (GIIN). The GIIN is a non-profit organization funded by the Rockefeller Foundation, J.P. Morgan, and the US Agency for International Development (USAID).⁴³ It was incubated by the Rockefeller Foundation and launched by the Clinton Global Initiative in 2009 with the goal of developing a collective voice for articulating and addressing the needs of the emerging impact investing industry: "The GIIN's goal is to share information on what works and what does not, to agree on common language and measures of performance, and to lobby for helpful laws and regulations."⁴⁴

The GIIN's Investment Council brings together about 40 institutions involved in impact investing and is made up of diversified financial institutions (from mainstream entities like J.P. Morgan and Prudential to development financiers like the Overseas Private Investment Corporation OPIC), institutional investors, specialized banks, targeted impact funds, and private foundations, major banks, institutional asset managers, and large-scale family offices.⁴⁵ "The Investors' Council is an *exclusive forum* for dialogue and activities to improve the effectiveness and efficiency of impact investing. Because these conversations are most productive when participants share common primary objectives, *only principal investors and capitalized investment funds are eligible for membership in the Investors' Council*"⁴⁶ (emphases added).

⁴³ <<http://www.globalimpactinvestingnetwork.org/cgi-bin/iowa/aboutus/supporter/index.html>>.

⁴⁴ See A Place in Society (n. 24).

⁴⁵ "The GIIN Investors' Council currently has 30 members: ACCION, Acumen Fund, The Annie E. Casey Foundation, Armonia (Lunt Family Office), Athena Capital, The Bill and Melinda Gates Foundation, Calvert Foundation, Capricorn Investment Group, CDC, Citi Foundation, Deutsche Bank, DOEN Foundation, Equilibrium Capital, FMO, Ford Foundation, The Gatsby Charitable Foundation, Generation Investment Management, Gray Ghost Ventures, J.P. Morgan, IGNIA, Leapfrog Investments, Lundin For Africa, Morgan Stanley, National Community Investment Fund, Omidyar Network, Overseas Private Investment Corporation, Packard Foundation, Prudential, The Rockefeller Foundation, Root Capital, RSF Social Finance, Saron Asset Management, ShoreBank International, Skoll Foundation, SNS Impact Investing, TIAA-CREF, The Tony Elumelu Foundation, Trans-Century, Triodos Investment Management, UBS, W.K. Kellogg Foundation," <<http://www.thegiin.org/cgi-bin/iowa/council/member/index.html#4>>.

⁴⁶ <http://www.globalimpactinvestingnetwork.org/cgi-bin/iowa/council/member_process/index.html>. Depending on their institutional size and type, Investment Council members make annual contributions ranging from \$5,000 to \$20,000. To determine membership eligibility, the GIIN considers the following criteria: (1) Does the investor currently manage assets committed to impact investment strategies? (2) In the case of investment funds, has the fund achieved a first close and begun investing for targeted social or environmental impact? (3) Does the investor strive to measure and report on the social and environmental impact of their investments? and (4) Does the investor prioritize participation in industry-building and collaboration with other investors in addition to pursuing their own individual objectives? (ibid).

Thus the GIIN, which is tasked with growing and institutionalizing impact investing through the production of industry measurement tools, such as indicators, and through the promotion of industry-friendly policy and regulation, has an expressly exclusive membership, which is today composed largely of North American and European financing entities. In light of this, and in light of its geographic and capital-mobilizing ambitions, it is easy to imagine how impact investing could be critiqued as a(nother) Northern idea with global aspirations that could come to have significant influence on the distribution of capital and of social and environmental goods across the global South.

A new framework for impact: IRIS and GIIRS

Two threads make possible the weaving of impact investing into a tapestry: IRIS and GIIRS. These recently launched initiatives provide key supports for establishing impact investing as a new asset class.⁴⁷ IRIS develops definitions, metrics, and industry standards for assessing social and environmental performance. The freely accessible metrics can be used by impact investing stakeholders to report on and track performance. IRIS is designed to provide an “independent and credible set of performance indicators that any investor, intermediary, or enterprise can use to measure, track, and report on organization’s social, environmental, and financial performance.”⁴⁸ GIIRS provides ratings services for a fee to companies receiving impact investments, as well as to the funds making impact investments. The ratings are made up of a numerical score, as well as a star score. The scores are based on the rating-seeking entity’s verified responses to a tailored survey.

While they are legally and substantively separate, GIIN, IRIS, and GIIRS are inter-connected: IRIS is housed within the GIIN (which in turn has support, including a space grant, from the Rockefeller Foundation) and fits squarely within the GIIN’s industry-building mandate. Additionally, the GIIN is an important vehicle for disseminating IRIS and GIIRS among Investment Council members. Moreover, GIIRS is a wholly owned subsidiary of a non-profit called B Lab, which has been heavily involved in developing the IRIS taxonomy. Finally, IRIS and GIIRS share funders, further justification that they should be considered together even though they serve different functions within the impact investing industry.

The central difference between GIIRS and IRIS is that IRIS does not judge performance; it does not rank companies or funds. In contrast, the central function of GIIRS is to judge performance—by issuing impact ratings—and to provide the investment and investee community with a precise tool for comparing companies and funds on the basis of their social and environmental impact. In this respect, IRIS does not meet the definition of “indicator” as laid out in the introductory chapter of this book, “Global Governance by Indicators.” That definition requires

⁴⁷ Emerging Asset Class (n. 1), 5.

⁴⁸ Amit Bouri, “How Standards Emerge: The Role Investor Leadership in Realizing the Potential of IRIS,” *Innovations*, Special Edition for SOCAP11, September 2011, 146 (hereinafter *How Standards Emerge*).

that indicators have certain features, including that they be ordinal, or capable of establishing a structure that enables comparison and ranking.⁴⁹ They should also generate “pressure for improvement, as measured by the indicator.”⁵⁰ As shown below, though it is not ordinal, IRIS can, much like an indicator, exert pressure to improve performance (or increase social and environmental impact), as well as influence decision-making and behavior, albeit with more room for maneuver.

In the creation of a new impact investing asset class, IRIS and GIIRS are vital producers of “market intelligence”:⁵¹ they generate meaningful and necessary information about the impact investing market as a whole and about the competitiveness of specific participants. They further serve to model the market for impact investing on the market for conventional commercial investing: IRIS is described as the impact investing equivalent of the International Financial Reporting Standards (IFRS) or the Generally Accepted Accounting Principles (GAAP), while GIIRS is described as the impact-equivalent of a Standard and Poor’s or a Morningstar ranking.⁵²

This packaging technique is deliberate; it is intended to attract investment through the strategic use of language and tools that are familiar to conventional commercial investors—scores and visual indicators, like stars. It is further intended to help quell concerns about the risks involved with channeling funds into a new investment area: “Imagine a commercial investing world in which there aren’t any rating agencies, or quantitative or qualitative risk measures: there would be no money coming into this world.”⁵³ As expressed by IRIS Director, Sarah Gelfand, the point of developing these tools is to build a market for impact investing that can attract conventional commercial investors: “This is absolutely meant to build a market for the for-profit investing world to participate, and to use standards, auditing practices, and rating agencies that would make [impact investing] more accessible to the broader investment community.”⁵⁴

Through their production of market intelligence, GIIRS and IRIS supply important informational links for guiding decision-making along the investor/investee chain. One of IRIS’s functions is to supply a glossary or dictionary for impact investing that equips investors and investees with the vocabulary necessary for transacting in this new area of activity. Another function, which is carried out via the production of impact metrics, is to clearly articulate the relationship between business activity and social/environmental performance. In this way, the IRIS metrics can be said to supply the grammar of impact investing. Thus IRIS, by

⁴⁹ Global Governance by Indicators (n. 3), 6.

⁵⁰ Global Governance by Indicators (n. 3), 7.

⁵¹ How Standards Emerge (n. 48), 156.

⁵² GIIRS’s stated purpose is to “assess the social and environmental impact (but not the financial performance) of companies and funds using a ratings approach analogous to Morningstar investment rankings or S&P credit risk ratings,” available at <<http://www.giirs.org/about-giirs/about>>.

⁵³ Monitor Report (n. 1), 66, citing Dan Letendre, Managing Director of the Merrill Lynch Community Development Company (in 2008).

⁵⁴ Tom Stabile, “Architects of a ‘Social Investment Data Engine’,” *Financial Times*, April 12, 2010 (hereinafter Stabile FT); see also Margot Brandenburg and Sarah Gelfand, “Impact Reporting and Investment Standards—Comment Now!” *Next Billion*, November 2, 2009 (hereinafter Next Billion).

producing both the vocabulary and grammar for impact investing, is establishing a common language for the industry.

The aggregation of metrics into a company or fund rating by GIIRS creates a concrete basis for impact comparison. This measurement component is considered essential for the growth of the industry and for un-shackling private capital that has been sidelined due to insufficient data and inadequate industry infrastructure. As explained by Gelfand, the target audience for IRIS and GIIRS are “the masses who would be willing to choose investments based on their social benefit if only they had a credible way to measure it.”⁵⁵ Industry leaders feel strongly that these tools, which communicate (primarily) non-financial information in a digestible fashion, are crucial for unleashing private capital into the social and environmental problem-solving sphere.

Such unleashing happens (in part) by lowering transaction costs:

GIIRS is driven by the vision of a scaled up impact investing marketplace with reduced barriers to entry and lower transaction costs. It is not the expertise of most investors to conduct an in-depth social or environmental performance analysis. Those that do have this expertise struggle to be able to compare the impact of their investments with investments outside of their impact investing portfolio. GIIRS ratings will make it easier for new investors to enter the impact investing marketplace by removing the need for an investor to have someone with social and environmental performance analysis skills on staff.⁵⁶

In other words, GIIRS can take on a share of the due diligence work usually done by firms, and also reduce the costs of carrying out due diligence by providing centralized and ready access to relevant information.

IRIS also carries transaction cost benefits: on the investor side, having shared definitions for impact can greatly simplify the process of evaluating different companies' information. On the investee side, reporting using one set of standards and definitions, as opposed to multiple standards required by different investors and donors, can save staff time and data collection resources and ultimately stimulate productivity.

Transparency is another necessary market feature that IRIS and GIIRS support and enable. Prior to issuing a rating, GIIRS requires entities to respond to a set of non-elective questions and ensures that answers are verified by a third party; this is intended to produce “comparable and transparent judgments about what is ‘good’ or ‘good enough’ in terms of a company’s social or environmental performance.”⁵⁷ Indeed, an important function of the IRIS/GIIRS impact framework is to enhance the credibility of impact investing by generating mechanisms for distinguishing “the good story tellers from the good problem solvers”⁵⁸ and for identifying

⁵⁵ Stabile FT (n. 54).

⁵⁶ Andrew Kassoy and Beth Richardson, “Who is the Fairest of them All? Creating a Rating System to Compare Social and Environmental Performance,” *Beyond Profit*, April–June 2010, available at <http://www.giirs.org/storage/documents/Beyond_Profit-4.10.pdf> (hereinafter *Who is the Fairest of them All?*).

⁵⁷ *Who is the Fairest of Them All?* (n. 56).

⁵⁸ MaRS (n. 12).

“investments that truly deliver double or triple bottom-line results versus those that just pay lip service to the social bottom line.”⁵⁹ Thus the framework aims to filter out red-washers or green-washers that claim to generate social and environmental benefits without substantiation. Such attention to transparency in turn serves to build investor confidence.⁶⁰

Because it quantifies social and environmental impact, transforming it into a measurable and comparable unit of value that can be considered in the calculation of the bottom line, the IRIS/GIIRS impact assessment framework can fundamentally change the way investors conceptualize and calculate returns on investments. The framework therefore alters the parameters and possibilities for investor decision-making. It also alters the context in which investees (actual and potential) make decisions about how to structure their business in order to maximize the particular values recognized by IRIS and GIIRS.

IRIS: Developing a common language for impact investing

The IRIS initiative was launched in 2009 by the Rockefeller Foundation, the Acumen Fund⁶¹ (a non-profit that invests in social enterprises in India, Kenya, and Pakistan), and by B Lab (a non-profit organization that issues social and environmental impact certifications for US companies while promoting the statutory recognition of hybrid corporations called Benefit Corporations).⁶² IRIS also receives support from professional service firms including Deloitte and PricewaterhouseCoopers. A public version of IRIS was launched in Fall 2010, and version 2.2 was released in the fourth quarter of 2011.⁶³ It is described as “a universal language for social, environmental, and financial performance reporting by mission-driven organizations.”⁶⁴

The IRIS initiative encompasses three components: (1) developing and refining the IRIS standards and metrics; (2) promoting adoption of the IRIS standards and metrics; and (3) enabling voluntary contribution of anonymous IRIS performance data to establish an evidence base of the industry’s performance.⁶⁵ To achieve the first component, IRIS provides a freely accessible menu of metrics for use by different types of entities (for-profit and non-profit businesses, as well as intermediaries, such as investment funds) for them to measure, track, and report on social, environmental, and financial performance. Users can select which IRIS metrics to track. The metrics menu takes the form of questions (approximately 430 can be selected) that break down into three main sections: (1) financial, which includes questions on revenue, assets, and profit margins; (2) operational, which

⁵⁹ Lisa A. Hagerman and Janneke Ratcliffe, “Increasing Access to Capital: Could Better Measurement of Social and Environmental Outcomes Entice More Institutional Investment Capital into Underserved Communities?” *Community Development Investment Review, Federal Reserve Bank of San Francisco* 5:2 (2009): 61.

⁶⁰ See, e.g., Media Advisory (n. 42).

⁶¹ <<http://www.acumenfund.org/about-us/about-us.html>>.

⁶² <http://www.giirs.org/assets/pdf/GIIRS-Sample_Rating2010_Complete.pdf>.

⁶³ <<http://iris.thegiin.org/iris-standards>>.

⁶⁴ Ibid.

⁶⁵ Ibid.

includes questions on internal governance, employees, wages, target customers/clients, suppliers, and emissions; and (3) sector-specific, which includes specific questions for entities engaged in agricultural production, the energy, environment and water sectors, education, community development finance, health care, and microfinance.

The questions are answerable in yes or no form (e.g., "Indicate whether the organization has a written child labor policy in line with International Labor Organization (ILO) standards"), or with a number (e.g., "Number of units produced"), or by checking a category (e.g., "Demographic groups of beneficiaries targeted by the organization's operations. Choose all that apply: Children and Adolescents, Disabled, Minority/Previously Excluded populations, Women, Other") or with a percentage (such as, "Growth in target population served"). With the exception of the mission statement and the description of the products and services provided, very few questions are answerable in prose form.

IRIS-compliant entities can anonymously report their metrics into a common data repository or "aggregator" that produces industry benchmarks and serves as "a repository to compile the data points on multitudes of organizations for market comparisons, benchmarking, and other analytics."⁶⁶ As more entities adopt IRIS and data submission increases, these benchmarks will allow investors to better contextualize the social and environmental performance of a potential investee, and give them the tools to assess how such investees fare against the industry standards. Conversely, an entity seeking investments will be able to measure its own performance against the industry benchmarks, and adjust its business strategy accordingly.

As mentioned, IRIS also serves a language-setting function by providing a glossary of definitions for the terms included in its metrics. The aim is for this taxonomy to be adopted throughout the impact investing industry in order to standardize (internal and external) reporting practices and establish shared definitions of terms used for assessing impact. Thus, in order to reduce transaction costs and avoid entities assessing performance using different, self-generated, or non-comparable criteria, IRIS sets out definitions of relevant terms, such as "job," "target population," "local supplier," or "small-to-medium enterprises," that can be used across companies, sectors, and geographies. The definitions serve a vital purpose because "until reporting is standardized with universal definitions, no reliable statements can be made about, say, the average number of jobs created by African healthcare startups, nor can fair comparisons be made between the social value of these organizations and, for example, similar new businesses in South America. The absence of definitional consistency renders performance metrics confusing to investors at best, and at worst, meaningless."⁶⁷

IRIS thus serves multiple purposes: it provides metrics for social and environmental performance tracking and assessment, a taxonomy for facilitating

⁶⁶ <<http://iris-standards.org/faq> and Stabile FT>.

⁶⁷ Sarah Gelfand, "How Social is Your Business: Creating a Standard Measurement System," *Beyond Profit*, January–March 2010, available at <<http://beyondprofit.com/how-social-is-your-business-creating-a-standard-measurement-system>> (hereinafter *How Social is Your Business?*).

comparable and harmonized assessments, and a system for defining benchmarks for the impact investing industry as a whole. Though IRIS does not rank companies, its benchmarking function facilitates comparisons between companies and between funds, which can influence decision-making and behavior. Companies that perform poorly with respect to IRIS-generated benchmarks could alter their practices in order to share more positive results in prospectuses and attract additional funding from new investors. Entities can also rely on (selected) IRIS metrics to report to their existing investors, governing bodies, or other stakeholders. Reporting entities could alter their practices in order to deliver more IRIS metrics-positive reports. In other words, when used for comparative or reporting purposes, IRIS metrics can acquire indicator features, particularly in terms of influencing behavior and exerting pressure for improvement.

A note on production

As concerns the production process for the IRIS taxonomy and metrics, the methodology is both consensus-driven and transparent. Its creators emphasize that it is a freely accessible and participatory public good, "IRIS must be developed as a non-proprietary public good if it is to achieve widespread adoption and overall success."⁶⁸ The current metrics menu was developed through the completion of both primary research (i.e., interviews and discussions with various experienced individuals and organizations) and secondary research (reviewing impact investing reports and existing taxonomies). The resulting indicators are an amalgamation and adaptation of metrics developed and used by other organizations, in addition to new metrics developed specifically for IRIS.⁶⁹

While the IRIS initiative is led by US-based, mostly private actors, it is also the product of a public feedback process: Anyone can visit the IRIS website and comment on any of the definitions or metrics listed, as well as suggest new metrics at <<http://iris-standards.org/framework/0001>>. The metrics can be viewed by genre (common descriptors, financial indicators, operations indicators, and sector-specific indicators). They can also be viewed by sub-section within a genre, meaning that one can opt to see, for example, only the indicators pertaining to water within the operations bracket. Comments provided through the website are "consolidated and the indicators refined to reflect the consensus opinions voiced through the Web site dialogue."⁷⁰

GIIRS: A ratings system for impact investing

GIIRS, a non-profit initiative, was created in 2009 and is a wholly owned subsidiary of B Lab, also a non-profit. B Lab previously designed a system for certifying US-based corporations as B Corps or Benefit Corporations, which has been used to

⁶⁸ Next Billion (n. 54); see also the Rockefeller Impact Investing Initiative Memo, available at <<http://www.rockefellerfoundation.org/uploads/files/774a2dac-af08-4ad1-b08e-95916952d6dd.pdf>>.

⁶⁹ <<http://iris-standards.org/faq>>.

⁷⁰ <<http://iris-standards.org/faq>>.

certify about 500 companies to date.⁷¹ GIIRS adapts the B Impact Assessment certification tool to rate domestic and international for-profit companies and investment funds on their social and environmental impact.⁷² The GIIRS tool was developed, tested, and revised during 2010 and 2011, culminating in an official launch at the Clinton Global Initiative in September 2011.

The ratings take the form of a star score—companies and funds can receive from one to five—and a percentage score, based on a maximum score achievable for a particular set of questions. In addition to the rating, GIIRS generates an impact report for each rating-seeking entity, which compares the social and environmental performance of the rated entity with that of its peers. In this way, GIIRS, like IRIS, generates benchmarks for use by impact investors and impact investees. GIIRS can be used by companies of different sizes (from 0 employees to more than 250), though there is an emphasis on SMEs. For the time being the ratings are not intended for publicly traded companies, for companies that have earnings of \$1 billion or more, or for non-profit entities.⁷³ Nor are they intended for public material funds, venture philosophy funds whose investments are mainly in non-profits, or funds that invest in real estate.^{73a}

Ratings are reached by weighing and aggregating the answers to 80–120 (mostly) IRIS-compliant questions contained in the Impact Assessment survey. Like IRIS, most GIIRS questions do not require a prose answer, and for those that do, the answers are not weighted, meaning that they are informational and not figured into the overall score. The number and content of the questions depends on a few key factors: whether the company is based in a developed country or in an emerging economy; the size of the company; the sector (service, wholesale/retail, manufacturing, and agriculture) it is engaged in; and the product or service (e.g., human health and social work, construction, and forestry and fishing) delivered by the company.⁷⁴

The questions relate to (1) governance (mission statement and financial accountability mechanisms); (2) workers (employee benefits, worker ownership, and employee policies and practices); (3) community (e.g., employment diversity, diversity of consumers/clients served, social and environmental screens for suppliers, charitable and civic engagement); (4) environment (e.g., environmental policies applicable to land and offices, sustainability of inputs, and product certification mechanisms); and (5) socially and environmentally-focused business models (e.g., models that incorporate environmental and social objectives into their governance structure, such as social enterprises; worker business models that emphasize worker ownership; community business models that are designed to meet the needs of a particular community; or environmental business models that are designed to preserve the environment).

As mentioned, industry leaders seem to be relatively agnostic concerning the motivations for making impact investments or for seeking a GIIRS rating. The

⁷¹ <<http://www.bcorporation.net>>.

⁷² <<http://bcorporation.net/resources/bcorp/documents/GIIRS%20USAID%20Press%20Release%204-26.pdf>>.

⁷³ <<http://www.giirs.org/companies/get-rated-companies>>.

^{73a} <<http://giirs.org/for-funds/get-rated-funds>>.

⁷⁴ <<http://www.giirs.org/companies/get-rated-companies>>.

assumption is that if an investor uses a GIIRS rating in making investment decisions, at least part of the intention is to achieve “impact”, even if a more important goal is to make a financial return or to win public relations points.⁷⁵ What does matter is having a solid governance structure in place to protect the ratings from the influence of those whose interests might compromise the quality of the rating.⁷⁶ As a result, the GIIRS business model is structured to ensure that ratings remain independent of the rated entities (a problem that plagued S&P, for example).

GIIRS assesses rating-seeking entities for a set fee, based on the entity’s annual revenue. The fee structure is intended to spread the costs associated with generating ratings across the rated companies and funds, and in particular to cover the expenses that GIIRS incurs when paying the third-party verifier of the information provided by ratings-seeking entities. Indeed, before ratings are issued, the information provided by ratings-seeking entities must be reviewed by a third-party verifier. Additionally, each year, 10 percent of the rated entities will be selected at random for on-site verification.⁷⁷ Currently, the preferred verification provider is Deloitte,⁷⁸ one of the GIIRS partners who participated in a \$6.5 million grant to B Lab to fund the development of the ratings system (along with USAID, Prudential Financial, and the Rockefeller Foundation).⁷⁹ Additionally, investors wanting to access the information underlying the ratings will pay varying subscription fees based on the breadth of information they seek. This dual payment stream model is designed to keep conflicts of interest in check by having both the entities seeking ratings and the investors seeking to access the ratings pay to use the system; this should ensure a high level of quality control.

A note on production

The B Impact Ratings System is governed by an independent Standards Advisory Committee made up of about 14 members drawn from a range of organizations including banks, pension funds, foundations, investor networks, ratings agencies, private wealth managers, and investment advisors.⁸⁰ Standards Advisory Councils (SACs) develop weightings and industry and region-specific standards for emerging markets (10 members) and developed markets⁸¹ (nine members).⁸² SAC membership

⁷⁵ Interview with Antony Bugg-Levine at the Rockefeller Foundation, September 2, 2010. Notes on file with author.

⁷⁶ Ibid.

⁷⁷ <<http://www.giirs.org/companies/get-rated-companies>>.

⁷⁸ Ibid.

⁷⁹ Media Advisory.

⁸⁰ <<http://www.giirs.org/about-giirs/giirs-team>>.

⁸¹ In defining when the Emerging Markets standards apply, the GIIRS team, including the SACs, consider different sources including the United Nations Human Development Index (HDI), the World Bank’s Doing Business Index, Gross National Income per Capita, Private Sector Credit/GDP Ratios, Transparency International’s Corruption Index, and UNDP’s Gender Empowerment Index. Initially, they only considered HDI and the Doing Business Index, but determined that this produced an incomplete and non-comprehensive picture of the developing world and expanded their sources accordingly. Interview with GIIRS staffer. Notes on file with author.

⁸² <<http://www.giirs.org/about-giirs/giirs-team>>.

is composed of a wide range of experts from non-profits, academia, and from the fields of performance evaluation, social finance, and international development, among others. Notably, as with IRIS, there is little representation by government (other than USAID), or by companies based in developing countries (as opposed to investors investing in developing countries), or by the clients or beneficiaries of impact investees.⁸³ The governance of GIIRS is designed to ensure that decisions concerning the content and weightings of the questions contained in the survey are made on a consensus-driven basis amongst the members of Standards Advisory Committee and the SACs. The SACs speak quarterly as a group, either by phone or in person, and members of the GIIRS staff keep the SACs updated on how GIIRS is being received on the ground and on any emerging challenges.

To prepare GIIRS for its official launch, B Lab organized a private beta phase, during which it tasked 25 “GIIRS Pioneer Funds,” representing leading venture capital funds with \$1.2 billion in assets under management and investments in 30 countries, and the companies that these funds have invested in, with testing the ratings system and providing feedback.⁸⁴ Two hundred companies were put through the beta in February and March 2011.^{84a} Members of the B Lab staff travelled to nine countries, and collected feedback from the fund managers of 13 emerging market investment funds and 135 of their portfolio companies to gather feedback—which, according to B Lab staff members, was largely enthusiastic and positive.⁸⁵

Like IRIS, the GIIRS production process is intended to be iterative, meaning that revisions will be made on a regular basis in order to ensure that user concerns, feedback, and needs are reflected. The GIIRS interface, like IRIS’s, makes it possible to provide feedback. However the GIIRS feedback process is considerably more burdensome than for IRIS: rather than having immediate access to all of the questions/metrics, as is the case with IRIS, people wishing to comment without seeking a rating are only able to access the GIIRS surveys by “pretending” to be a company (free of charge), and setting the parameters (geographic location, sector, industry, number of employees, etc.) that will then generate a tailored survey.⁸⁶ Once “in” the survey, feedback buttons appear after each question, allowing the submission of comments. The feedback buttons will remain active throughout the life of the tool and will allow revisions and updates to be made on a large scale once every two years, if not more frequently. It is not clear at this stage whether the feedback and comments recorded through this process will be made publicly available, as is the case with IRIS.

⁸³ Developed Markets: B Lab, Risk Metrics, Workplace Dynamics, Duke University Center for the Advancement of Social Entrepreneurship, Skoll Foundation, GoodGuide, SJF Advisory Services, Pacific community Ventures, RSF Social Finance. Emerging Markets: Aspen Network of Development Entrepreneurs, Dalberg Global Advisors, World Resources Institute, Grassroots Business Fund, Technoserve Kenya, US Agency for International Development, Acumen Fund, B Lab, Harvard University Center for Business and Government, and GIIN.

⁸⁴ Q1 Progress Report.

^{84a} Q1 Progress Report.

⁸⁵ Interview with B Lab staff member. Notes on file with author.

⁸⁶ Note that it is quite difficult to change the parameters once they are set.

Returning to the central difference between IRIS and GIIRS, this sub-section has shown that IRIS can “graduate” from metric to indicator when used by stakeholders to compare and/or change social and environmental performance, as in accordance with IRIS-generated industry benchmarks. In this sense we suggest that while GIIRS is an indicator proper, IRIS is an indicator-based system. This is not to say that the status of indicators is superior or preferable to that of metrics. In fact, as a tool for tracking and improving performance along social and environmental dimensions, IRIS may be more useful than GIIRS, precisely because it does not rely on a pre-set formula for assessing impact. Indeed, by allowing users to select which metrics to track, and tailor their selection to specific interests and needs, IRIS creates more room for maneuver and is perhaps better suited to the sectoral diversity and novelty of impact investing than GIIRS. Additionally because it does not provide packaged company judgments, it may be easier to combine IRIS with other performance assessment tools that are more qualitative.⁸⁷

On the other hand, the fact that GIIRS does not permit rating-seeking entities to select the criteria on which they will be judged sets a concrete, and possibly more objective reference point for comparison: Every company falling into a particular geographic, sector, and size combination will be judged the same way. For investors with a clearly defined interest, and perhaps more limited due diligence capacity, GIIRS is a one-stop shop: GIIRS selects which IRIS metrics to track on the investors’ and investees’ behalf, assigns weightings to answers, and aggregates them to produce a rating. In contrast, IRIS seems more diffuse, precisely because it is more modular, and because it leaves more responsibility in the hands of users than GIIRS. Recall that the aggregator function provided by IRIS can only go some of the distance toward the one-stop shop model; this is partly because the aggregator is not designed to generate company judgments—only industry benchmarks—and partly because it is entirely dependent on *voluntary* data contributions. The utility of the IRIS aggregator may ultimately be limited, therefore, as it is the users who determine what type and volume of data to submit.

IRIS and GIIRS perform different functions within the impact investing industry. They can be used to tell the same impact story differently, or even to tell different impact stories altogether. They are complements, not substitutes, and this is why they are treated together here as an impact assessment framework.

Challenges to the market-building governance function of indicators

The previous sub-sections showed that the GIIRS/IRIS impact assessment framework is essential to the project of building a market for impact investments; it further explained how the framework could influence the behavior of both investors and investees by shaping and informing investors’ decision-making and the selection of particular business models and strategies by investees, for example. In

⁸⁷ How Standards Emerge (n. 48), 150.

its market building function and in its role as informant and guide for decision-making within the industry, the framework exercises a governance function. As a result, it may face certain challenges going forward. Such challenges relate to the infrastructural readiness of impact investing for attracting and channeling investments on the one hand and to the legitimacy of the impact assessment framework on the other. Indeed, should this framework comes to be an ordering mechanism for impact investing, channeling capital into and within the industry, then stakeholders who are disadvantaged by it, as well as actors with competing claims to the authority of IRIS and GIIRS, may contest the framework's legitimacy.

On infrastructural readiness: as mentioned above, average impact investments are currently estimated at \$1 million;⁸⁸ this amount may be too low for a number of the bigger investors, even though the latter are considered essential for mobilizing the capital necessary to correct the deficit in the global social balance sheet. Indeed, the transaction costs involved with investing in 30 small projects are much higher than investing in a smaller number of larger projects, and generally, serving the poor is relatively expensive, which means that "there is often a need to invent new and disruptive business models" in order to make business propositions palatable to big investors and effective from the point of view of smaller investees.⁸⁹ Yet, as some observe, this type of invention is something that "existing players lack an incentive to do."⁹⁰ Thus, at this stage, the business models and financial intermediary structures in place may not be either developed or diverse enough to attract and retain the sought-after bigger investors.

A related concern is that the capital absorption capacity of the businesses currently operating in the impact investing sphere is too limited to process large amounts of financing and achieve meaningful social and environmental impacts. In other words, capacity building at the investee level may be the most pressing agenda item, not lowering the cost or increasing the volume of capital. Should incapacity at the investor or investee level materialize into stumbling blocks for the impact investing industry, IRIS and GIIRS risk being criticized for being ineffective at connecting investors and investees.

Worse, if too much capital begins to chase firms that are unable to use it productively, an impact investing bubble could be created, a bubble that would burst when firms begin to report disappointing returns. This too could result in industry failure, with investors not receiving expected or needed returns on investments and abandoning the industry altogether, which would produce negative consequences for investees and their client base. GIIRS and IRIS, however well designed, might be viewed as premature if the infrastructure for supporting impact investing is inadequate; this, in turn would impair their governance function. This generally points to the importance of having a critical mass of buy-in for the impact investing proposition and its tools.⁹¹ Time will tell whether the current buy-in levels are sufficient to weather any serious infrastructure deficiencies.

⁸⁸ Emerging Asset Class (n. 1), 35.

⁸⁹ Monitor Report (n. 1), 22.

⁹⁰ Ibid.

⁹¹ See *How Standards Emerge*, 146, on the need for a critical mass of early adopters to find IRIS's value proposition compelling at an individual level in order for it to become the de facto standard for the impact investing industry.

Further, if, as predicted by industry leaders, deal size increases over time, some challenging questions arise: would smaller projects be left out because they are not considered investable for larger investors or because it is easier to create and fund larger projects than to develop new financing mechanisms and (riskier) business models? Relatedly, and this is more fundamental to the integrity of the project, what if the firms whose operations have the greatest social or environmental impact do not have correspondingly high needs for capital? Would an industry focus on large deals, to the detriment of smaller transactions or capacity building, run counter to the goal of correcting the deficit in the global social balance sheet, limiting the positive social and environmental outcomes? Such infrastructural concerns threaten the legitimacy of IRIS and GIIRS.

Another issue affecting GIIRS in particular is that while it may lower the barrier to entry for investors, it might also raise the barrier to entry for investees. Though GIIRS is intended for use by small and medium-sized companies, and though the cost of obtaining a rating is proportionate to a company's annual revenues, it might still exclude smaller entities that lack the human and logistical resources and capacity to partake in the ratings process. And even if small entities do partake, they may not get a good rating for reasons that are beyond their control. As an example, a company may get a low score because it procures materials from a supplier with a poor CSR record; but, depending on the setting, finding a "good" supplier may be challenging or impossible, particularly in poorer countries. Conversely, using an input like scrap metal in production, which is tantamount to recycling, could increase a score if the metrics are sufficiently sensitive; if not, however, such green processes may not receive the rating they deserve, which may even create incentives for adopting counter-practices.

A second group of challenges relates to competitiveness within the impact assessment realm. GIIRS and IRIS are not the only tools for assessing impact (see the many alternative approaches to assessing social impact listed on the website, TRASI or Tools and Resources for Assessing Social Impact⁹²). In fact, they explicitly draw inspiration from other approaches.⁹³ They are unique in that they are tailored to assess impact *across* geographies and sectors—water, education, health, etc.—as distinct from, say, the CERES Principles, which are focused on environmental standards, reporting, and certification, or the Microfinance Information Exchange (MIX) reporting tools, which are tailored for the microfinance sector. Further, they extend not to one, or two, or even 15 projects or companies, but potentially to thousands, whereas many of the impact assessment tools available are designed to produce information only about a particular project or set of projects, without generating comparable results.

The central value-added of the GIIRS/IRIS impact assessment framework is its potential for institutionalizing, in a precise and relatively versatile fashion, the

⁹² <<http://trasi.foundationcenter.org>>.

⁹³ See "The Global Impact Investing Network (GIIN): An Interview with CEO Luther Ragin, Jr," *Forbes Magazine*, October 30, 2011, describing IRIS as a "meta-standard" that incorporates existing performance reporting standards where available.

CSR-plus approach described above by incorporating impact, in addition to harm-avoidance, into investor and investee decision-making. However, even along this dimension, GIIRS and IRIS do not have a monopoly, though they may have a head-start. Other systems, such as the Global Reporting Initiative (GRI), which has drawn criticism for “lacking teeth,” but which has the advantage of being well-known and quite widely used, seem already to be expanding to incorporate impact (or something similar).⁹⁴ GRI is viewed as being better suited to large companies with extensive supply chains, and so may only be a minor threat to the IRIS/GIIRS impact framework; however, if the impact investing industry broadens its corporate demographic to include larger investee companies, or if GIIRS begins to provide ratings to publicly traded companies, then the GRI might stake its claim as a better performer of governance functions.

Similarly, it could be argued that, rather than operating independently, the IRIS standards should be incorporated into reporting mechanisms that are already being used by global businesses (and that IRIS emulates) such as the IFRS or the GAAP. Such a shift could potentially achieve more to improve the quality of social and environmental accounting and reporting on a large scale than introducing an entirely new system. Likewise, GIIRS could be more effective as a signaling device if it were integrated into an established ratings system, such as S&P or Moody's. And if the latter continue to expand to include social and environmental performance, then GIIRS may lose its niche.⁹⁵ Interestingly, there could be a causal connection between this second group of challenges and the first: if the infrastructure issues that threaten the legitimacy of IRIS and GIIRS are resolved, meaning that impact investing becomes a viable asset class and a prosperous industry, then competition will no doubt kick in, triggering the risk that IRIS and GIIRS will be displaced in their exercise of market-building governance functions by alternative systems for accessing impact.

Indicators and government regulation

Turning to governmental decision-makers, this section explores some of the ways in which, particularly where regulation is under-developed for purposes of overseeing a new area of activity and asset class, indicators can influence, complement, and even substitute for government-led intervention. It argues that where indicators serve a regulatory function, they should be evaluated accordingly, especially with respect to criteria such as legitimacy.

⁹⁴ See Rachel Sanderson, “Initiative to Overhaul Global Reporting Standards,” *Financial Times*, August 1, 2010. See also Marcy Murnighan, “The Three I’s of Sustainability Reporting’s Future,” 2010, available at <<http://www.globalreporting.org/NewsEventsPress/LatestNews/2010/TheThreeIsofSustainabilityReportingsFuture.htm>>.

⁹⁵ See, e.g., Moody's Press Release: Moody's to Develop Social Performance Assessment of Microfinance Institutions, September 22, 2010, available at (n. 1) <<http://murnighanpost.com/2010/11/08/the-three-is-of-sustainability-reporting%E2%80%99s-future>>; see also Harnessing Capital Markets (n. 1), 34, on the increased incorporation of social and environmental issues in financial performance assessments.

Regulating a new industry: Demands on government

With profit-driven and mission-driven moneys being deployed to address the same set of problems, the demarcations between commercial investment and philanthropic giving are beginning to blur.⁹⁶ In the US, this blurring has been accompanied by some regulatory innovation in the space that lies between but, as stakeholders often bemoan, this change has not kept up with developments in either the investment or the philanthropic realms. A recent report by Insight at Pacific Community Ventures and the Initiative for Responsible Investment at Harvard University entitled, "Impact Investing: A Framework for Policy Design and Analysis," aims to fill this gap by fleshing out the role that government policy can play in supporting the impact investing industry.⁹⁷ The report's publication and its content highlight that getting government on board to nurture the growth of impact investing is a high priority for industry proponents.

The report emphasizes the role of government as a key industry facilitator and explains that tailored policies can affect each of the "three pillars" of the impact investing market by (1) increasing the supply of capital from investors; (2) increasing the demand for impact-oriented capital from companies; and (3) directing capital toward impact investments at the point of exchange, where rules govern the terms of trade.⁹⁸ It highlights that while government can provide direct financial or institutional support for impact investing and influence markets through laws and regulation, the real effectiveness test for policy is whether it succeeds in "lowering barriers to, or otherwise catalyzing, investment for social and environmental benefit" and in creating "viable private markets that support the creation of social goods efficiently."⁹⁹ Further, because "[i]mpact investing has the potential to provide government with a powerful delivery mechanism for social goods,"¹⁰⁰ policy should be less focused on restrictively regulating impact investing, and more focused on facilitating the (independently increasing) delegation by government of its function as a provider of social goods to the private sector generally and to market-based systems more specifically.

In order for such delegation to be successful, new policies, "from regulations governing how investors deploy capital, to mandates for specific social and environmental provisions within markets, to support and subsidy for identified public goods"¹⁰¹ must be designed and implemented. The report recommends that this

⁹⁶ Note that some argue that the separation of philanthropy and commerce, dating back to somewhere around the 1700s was "a 'historical accident, a giant navigational error' and that 'the inertia of [the initial division] remains strong', resulting in modern charities being held back in their growth." Matthew F. Doeringer, "Fostering Social Enterprise: A Historical and International Analysis," *Duke Journal of Comparative & International Law* 20:2 (2010): 303, citing Bill Drayton, "The Citizen Sector Transformed," in Alex Nichols (ed.), *Social Entrepreneurship: New Models of Sustainable Social Change* (Oxford University Press, 2006), 45 (hereinafter *Fostering Social Enterprise*).

⁹⁷ Insight at Pacific Community Ventures & the Initiative for Responsible Investment at Harvard University, "Impact Investing: A Framework for Policy Design and Analysis," January 2011, Supported by the Rockefeller Foundation, available at <<http://www.rockefellerfoundation.org/news/publications/impact-investing-framework-policy>> (hereinafter *A Framework for Policy*).

⁹⁸ *Ibid.*, 18.

¹⁰⁰ *Ibid.*

⁹⁹ *A Framework for Policy* (n. 97), 11.

¹⁰¹ *Ibid.*, 12.

reform be undertaken with care, however: “We caution that impact investing policies, if they favor private market activity blindly, may undermine the delivery of public goods that governments are designed to provide . . . Fundamental considerations of equity and other national priorities should remain paramount in policy design.”¹⁰² Thus the report (and industry proponents, whose views it captures) puts dual pressure on governmental decision-makers: to serve and grow the impact investing market through new, enabling policies, on the one hand, and to police the market to avoid compromising the delivery of public goods, on the other. We will see that IRIS and GIIRS could have a significant role on both fronts.

What follows is an outline of the regulatory interventions and other decisions that government may be called upon to effect as impact investing is established as a new industry and asset class:

- (1) Increase incentives for channeling private capital into impact areas via measures such as tax credits, exemptions, and other subsidies.

Impact investing promoters seek government support to stimulate private investment. They suggest making low interest-rate loan instruments available, for instance, or leveraging the government’s capacity as a large-scale purchaser by “providing anchor demand for promising enterprises,” which could enable such enterprises to scale up their businesses.¹⁰³ Such measures would incentivize private investment through leveraging, but also through mitigating the risk/return tradeoff involved with entering a new industry where financial returns may be diminished because of the focus on generating social or environmental returns.¹⁰⁴ Additionally, as new rules continue to be developed in response to the financial crisis, exemptions will be sought for the industry. As an example, an exemption could be made for impact investments to the Volcker rule, which limits the ability of US banks to make speculative investments in private equity.

- (2) Increase and harmonize the statutory recognition of hybrid entities or mission-driven for-profits.

Hybrid entities, or mission-driven for-profits, are natural recipients of impact investments. Benefit Corporations, mentioned earlier, and Low Profit Limited Liability Companies (L3Cs) are examples of such hybrids. Both involve for-profit entities that incorporate double or triple bottom lines into their business models. These entities, which are sometimes described as “for-profits with a non-profit soul”, are able to access equity capital, which is something that registered non-profits cannot do without running afoul of non-profit legislation.¹⁰⁵

¹⁰² Ibid., 16.

¹⁰³ Monitor Report (n. 1), 69.

¹⁰⁴ See Monitor Report (n. 1), 68: such support was provided through the Community Reinvestment Act of 1977, which “was pivotal in creating the community development finance industry that has since poured billions into otherwise underserved neighborhoods.”

¹⁰⁵ See John Romano, “L3Cs: A Sweet Honey of a Deal for Non-Profits,” on the Socialenterpriseblog, available at <<http://socialenterpriseblog.com/?p=9>>: “A foundation looks at its PRI investment money to further the ‘effectiveness’ of its charitable dollars beyond making just an outright grant . . . Furthermore, the amount of PRI investment counts toward the mandatory 5 percent that foundations

The legislative status of these entities is decided at the state level. L3Cs can currently be formed in nine states and two Indian nations;¹⁰⁶ once formed, they can operate legally in all 50 states though the tax treatment and status of L3Cs is uncertain for the time being,¹⁰⁷ though L3Cs can receive program-related investments (PRI) from foundations, which allows foundations to make profit-generating investments that are related to their charitable program without compromising their private foundation status.¹⁰⁸ Seven states have passed Benefit Corporation legislation to date, and several more are considering doing so.¹⁰⁹

A major value-added of these forms relates to branding: entities incorporated as L3Cs or Benefit Corporations communicate unambiguously to investors, clients, and consumers that they are mission-driven and that they weigh social and environmental returns heavily. However powerful branding may be, the lack of statutory and fiscal recognition of these entities is viewed as a hurdle to establishing sustainable businesses that can achieve multiple bottom-line returns, and effect meaningful change in the ways business does business.¹¹⁰

Hybrids are also connected with a redefinition of corporate fiduciary duties. Indeed, as new corporate forms, Benefit Corporations and L3Cs could redefine conceptions of fiduciary duty, which is important for two reasons: First, companies can pursue a double or triple bottom line, and possibly compromise profits in doing so, without fearing shareholder lawsuits brought for breach of duty—on grounds of sacrificed financial returns. Second, building on the analogy frequently used by B Lab, a re-conceptualization of fiduciary duties could bake a new kind of social and environmental accountability into the DNA of corporate entities¹¹¹ if shareholders, instead of resisting diminished profits, come to demand increased social and/or environmental returns and hold management accountable for delivery.

are required to give out in order to maintain its legal status as a private foundation . . . The mission of an L3C is a social benefit, doing socially productive and useful things, and only then earning a profit.”

¹⁰⁶ <<http://www.americansforcommunitydevelopment.org/laws.php>>.

¹⁰⁷ See Malika Zouhali-Worrall, “For L3C Companies, Profit Isn’t the Point,” on *CNN Money.com*, February 9, 2010, available at <http://money.cnn.com/2010/02/08/smallbusiness/l3c_low_profit_companies>.

¹⁰⁸ For a detailed review, see *Fostering Social Enterprise* (n. 96), 316–21, explaining that production of income cannot be a significant purpose of the PRI investment and that any income received from investment in PRIs is not counted as profit for the foundation but must be reinvested that year.

¹⁰⁹ <<http://www.benefitcorp.net/state-by-state-legislative-status>>.

¹¹⁰ See, e.g., Antony Bugg-Levine, Interview, @lliance for Philanthropy and Social Investment Worldwide, May 1, 2010, available at <<http://www.alliancemagazine.org/node/3283>>.

¹¹¹ Major provisions of B Corp legislation include: a duty on behalf of directors and officers to consider the effect of decisions on stakeholders, not only shareholders; directors and officers will also have discretion to give priority to particular stakeholders consistent with general and public benefit purposes; And, on an annual basis, the corporation will publish a Benefit Report in accordance with recognized third-party standards for defining, reporting, and assessing social and environmental performance. <http://www.csrwire.com/press_releases/29332-Maryland-First-State-in-Union-to-Pass-Benefit-Corporation-Legislation>.

- (3) Develop regulations and standards for overseeing social and environmental finance that would stimulate investor confidence, enhance transparency, and help ensure that green- and red-washing are avoided.

Developing regulations and standards for social and environmental finance could prevent impact investing from growing too fast into an asset class that is unable to sustain itself, or the expectations it raises. It would help mitigate the risks that might materialize if, in spite of good intentions, the definition of social and environmental impact “turns out to be so loose and diluted as to be virtually meaningless,” which could have the unfortunate result of diverting capital away from philanthropy, thereby “decreasing the resources dedicated to addressing serious societal challenges.”¹¹²

Indeed, as mentioned earlier, one important serious concern is that the “hype” around impact investing, and the increasing reliance by citizens and government on market-based systems to provide and distribute social goods, could create an investment bubble; one that could burst if “there is a significant gap between expectations for financial and social returns and actual performance, which may happen if the concept is sold ahead of demonstrated social impact and/or economically viable deal flow.”¹¹³ In other words, policy and tailored rules could help to avoid a situation where “[p]oor thinking and sloppy execution” produce returns that are too low to meet expectations, and, more alarmingly, mis-channel valuable resources intended for addressing social challenges.¹¹⁴

- (4) Formulate consumer protection principles to nurture a healthy blended value world.

Consumer protection principles are inadequately developed in the impact investing context. In the commercial context, we know how to define consumers, shareholders, and their rights. The same can generally be said about donors and grantees in the philanthropy context. Consumers can vote with their wallets and generally seek protection from government if their rights are violated. What is the equivalent for the beneficiaries of programs financed through impact investments? Crucially, who *are* the consumers in a blended value world? What rights and redress should they have as billions of return-expecting dollars are mobilized to address social and environmental problems? It seems important to examine these questions closely, particularly as the production and distribution of (often vital) social goods are at issue.

Interim indicator measures

As illustrated above, government may be called upon to design various interventions and policies to support the growth of impact investing. There are several ways that IRIS and GIIRS might be relied upon in the course of making such design and

¹¹² Monitor Report (n. 1), 35.

¹¹³ Ibid.

¹¹⁴ Ibid. Note that Britain may provide a model here, having institutionalized a Community Interest Company (the British equivalent of a hybrid entity) Regulator to register and monitor compliance with regulations that are specific to mission-driven businesses.

implementation decisions, including by providing the basis, and even standing in for, official interventions.

One can easily imagine government relying on IRIS and GIIRS to determine which entities should be eligible for assistance, whether in the form of subsidies, tax exemptions, or other special treatment, for instance. Or they could be used to determine eligibility criteria for the incorporation of hybrids. As an example, through the survey and assurance process, GIIRS ratings attest to companies' social and environmental performance; this type of information could form part of the information package required for determining L3C, PRI, or Corporation Benefit eligibility. GIIRS could therefore serve as a due diligence tool (or even arm) for government. The same could be said for granting subsidies or tax benefits to mission-driven companies or funds: entities with high GIIRS ratings could be deemed presumptively eligible for certain benefits.

GIIRS and IRIS could also inform the reporting and disclosure requirements that might be applied to mission-driven businesses and impact investment funds for consumer protection purposes. As an example, the IFRS were initially adhered to on a voluntary basis and are currently required for all European Union-registered companies; they have also been recognized by the US Securities and Exchange Commission as acceptable substitutes for the GAAP (for non-US companies). It is conceivable that the IRIS reporting standards could follow a similar journey and be adopted by an existing regulatory entity, or possibly a new body created for the purpose of overseeing impact investing (as exists in Britain).¹¹⁵ This would constitute a direct expansion of official regulation via the adoption of privately generated standards.¹¹⁶

Additionally, through the taxonomy, industry benchmarks and standards, and ratings, GIIRS and IRIS can enhance the accuracy of the risk/return tradeoff assessment carried out by investors. Thus at this stage in the industry's development, it could be argued that the information generated by IRIS and GIIRS stands in for the information typically required and disseminated by government in the commercial investing context; it is therefore at least conceivable that such information could become incorporated into government practice down the line.

Until government assumes explicit regulatory functions within the impact investing sphere, the fact that both IRIS and GIIRS have financially and reputationally strong networks of supporters, and that they currently have no direct competition, lends credence to the view that stakeholders will come to rely on them as the primary rule-setters and governors of the social investing space. Even without official adoption, GIIRS and IRIS could exercise significant influence, by channeling funds toward "impactful" entities. If adopted by government, their governance function could expand and grow in proportion to what we refer to as the "market for impact".

¹¹⁵ Fostering Social Enterprise (n. 96), 312–13.

¹¹⁶ On IFRS, which was generated by a London-based non-governmental organization, and its role in expanding global regulation, see Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton: Princeton University Press, 2011).

Because of their function as interim regulators, IRIS and GIIRS may themselves invite regulation. Rather than being evaluated through votes or conventionally democratic procedures,¹¹⁷ these tools will, at least initially, be assessed through usage statistics, such as the number of companies inputting data into the IRIS aggregator or number of GIIRS subscribers, for instance. Their effectiveness as market-builders will likely also be assessed by looking at the amounts of capital that flow in reaction to the information they produce and disseminate. To some degree then, the framework has built-in quality control mechanisms, and it could be argued that it is self-regulating. Nevertheless, because IRIS and GIIRS can influence government behavior and decision-making, and because the governance functions exercised by IRIS and GIIRS extend to the production and distribution of social goods, a function that has traditionally been exercised by government, scrutiny such as that applied to public actors and official policy seems warranted.

Challenges to the regulatory governance function of indicators

Even early in their indicator-lives, GIIRS and IRIS illustrate how indicators can serve as incubators for regulatory intervention. They serve a governance function by charting a course, and even by serving as placeholders for official regulation. Arguably, they also serve as regulators in their own right. Such a view is supported to the extent that costs to “non-compliance” can be identified. Here, such costs would amount to not receiving a GIIRS rating, not participating in the benchmarking process, and possibly falling off (or never coming onto) the investor radar—and, for that matter, the government radar if GIIRS and IRIS come to be relied upon by policymakers. And if IRIS succeeds in becoming the common language for impact reporting, non-compliant or non-IRIS-speaking entities could find themselves reporting using vocabulary and standards that are no longer meaningful.¹¹⁸

GIIRS and IRIS embody judgments about what constitutes social and environmental impact and about the right way to assess such impact.¹¹⁹ Because these judgments can affect the financing, production, and allocation of social goods, they trigger public interests, which invites additional scrutiny. The same can be said for the governance functions exercised by IRIS and GIIRS, which, like their relationship to official regulation, are relatively complex. Drawing on governance concepts such as participation, effectiveness, accountability, and responsiveness, this subsection considers the legitimacy of the IRIS/GIIRS impact framework. In doing so,

¹¹⁷ See Davis, Kingsburg, and Merry, *Global Governance by Indicators* (n. 3), 76–8: “The scientific methods used to produce indicators tend to involve distinctive social processes . . . producing, accepting and supporting standards embedded in indicators engages social processes more typically with the derivation of power from scientific processes.”

¹¹⁸ See *How Social is Your Business* (n. 67), for a useful example from the microfinance arena of how the adoption of the Microfinance Information Exchange (MIX) metrics became so popular as to make alternatives almost irrelevant.

¹¹⁹ See generally, Davis, Kingsbury, and Merry, *Global Governance by Indicators*, (n. 3).

it also considers avenues for contestation. To begin, recall that the GIIRS/IRIS impact framework is intended to (1) attract more capital to solution-driven projects globally; (2) increase transparency to distinguish good story-tellers from good problem solvers; and (3) bring about change in the way that business does business to be stakeholder-oriented, rather than shareholder-oriented.

Participation

On whose behalf are IRIS and GIIRS, as interventions, being carried out, and who participates in their design? Given the explicit focus on mobilizing funds to solve social and environmental problems so as to correct the deficit in the world's social balance sheet (with a significant emphasis on emerging economies), it seems fair to describe the production of GIIRS and IRIS as being carried out on behalf of those whose lives can be improved through impact investing-funded projects. For convenience, this group will be referred to as "beneficiaries" from this point forward (bearing in mind the earlier discussion on the difficulty of identifying the intended beneficiaries of impact investing). Yet, as the governance structure of both IRIS and GIIRS shows, with predominantly US and North European-based private entities at the helm, these beneficiaries are not directly involved in the production of the framework. While the public is invited to participate in the online feedback and comment processes described earlier—setting aside the issue that many beneficiaries (especially in poor countries) may not have internet access to contribute—the weighting and definitional decisions are made without direct beneficiary input.

Further, neither IRIS nor GIIRS require any detail as to the substance of beneficiary or client feedback on impact investing-financed projects. That is, positive beneficiary feedback is not a criterion for a high GIIRS rating, nor is it a metric listed in the IRIS menu that can be tracked. While the consensus-driven production process of GIIRS and IRIS mitigates some of the risks involved with not including beneficiary feedback among the items being tracked and aggregated, it does not fully address the concern that the market for impact is being developed largely without substantive input from the "impact-*ed*." This could create avenues for contestation by beneficiaries, as well as by entities that may be very well viewed by their beneficiaries, but unable to see this reflected in a GIIRS rating. This leads to additional concerns about what is counted as "impact" and from whose perspective. It also touches on the challenge of capturing something as politically charged, subjective, and essentially participatory as social and environmental impact in the form of a metric or an indicator.

Responsiveness

It is useful to consider whose interests are reflected in the framing of social and environmental problems by IRIS and GIIRS, and whether stakeholders with different interests may have conflicting framings. GIIRS (and to a lesser extent IRIS) clearly prioritizes the financiers of market-based solutions because the latter can potentially deploy large amounts of capital in service of solving social and

environmental problems. Perhaps the content of “impact” would look different if the intervention were less centered on improving investor ability to identify and compare investments, and more on identifying the most pressing needs at the local level, from beneficiaries’ perspectives.

A related question is whether the purpose of the intervention is well served by the apparent agnosticism concerning the priorities of investors. In so far as IRIS and GIIRS create a new paradigm for defining, measuring, and valuing impact, and to the extent that their investor-agnosticism influences that paradigm, there is a risk that the complexity of beneficiaries’ needs will become detached from the impact equation. Indeed, perhaps the tools would be constructed differently and be more successful at attracting impact-capital and at securing government support if they addressed those investors who explicitly (and accountably) prioritize social over commercial returns. This again opens contestation channels for beneficiaries (actual and potential), as well as for the entities serving them.

As concerns stakeholders with potentially conflicting interests—namely investors, shareholders, and beneficiaries—possible tensions arise from the question whether GIIRS accords appropriate weights to the raw data collected for the purpose of assigning a rating. Indeed, no single weighting scheme is likely to satisfy all investor accords.¹²⁰ Some investors are more interested in profit, others are more interested in social impact; some are willing to support small or untested firms, while others are more interested in larger or more established entities; some will discount long-term impacts in favor of short-term results, whereas others might be more interested in supporting impacts that materialize only in the longer term; and of course, some will seek out interventions that focus on particular beneficiaries with particular needs. Thus an individual investor may discredit the framework on the basis that it inadequately captures the type of impact she seeks to achieve. Similarly, a beneficiary may find that there is a big discrepancy between a rating and their experience of impact. As a result of these varying interests, investors as well as beneficiaries, may view the framework as biased and illegitimate.

Accountability and transparency

The developers of the IRIS/GIIRS framework acknowledge a risk of mission-drift within the industry: “As purely commercial investors (that may not be committed to ‘double bottom line’ business) take stakes in impact investments, observers fear that the companies may succumb to pressure to prioritize financial returns over social impact.”¹²¹ Such concerns have strong resonance in the microfinance sector, where initial public offerings of microfinance institutions have elicited objections that the social mission will become only a secondary priority for new shareholders, and that the profitability of these institutions will come before the financial health

¹²⁰ See *Emerging Asset Class* (n. 1), 75, highlighting the subjective nature of social impact and the need to institute standard approaches to impact measurement to enhance transparency and objectivity in investor decision-making.

¹²¹ *Emerging Asset Class* (n. 1), 67.

of the generally poor borrower base.¹²² The risk of mission-drift may be aggravated by the limited involvement of beneficiaries and government in developing and overseeing the impact framework. To the extent that IRIS and GIIRS succeed in becoming norm-setters for impact investing, mission-drift poses a serious accountability risk, particularly as the losers would be the very beneficiaries whose lives impact investing seeks to improve and the causes that it aims to advance.

Industry advocates view good metrics as key transparency and confidence-building tools that help to distinguish good story-tellers from good problem solvers. They further “believe that mission drift (or even false claims of an impact mission) and exploitation are legitimate concerns” and that “impact investors should ensure that the right metrics are in place to monitor their portfolio companies.”¹²³ Thus advocates rely on metrics to produce transparency and do the mission-guarding work, but seem to sidestep the issue of how metrics themselves might contribute to the problem, which begs the question, what *is* the “right” metric. And if, especially given the potentially serious conflicts of interest among stakeholders, the metrics fail to guard missions, the losers, meaning the beneficiaries but also those investors who channeled funds into impact investments to produce social and environmental benefits, would have strong grounds on which to contest the legitimacy of both IRIS and GIIRS.

Efficacy

There is reason to question the accuracy of the results produced by the impact assessment framework. Specifically, it is unclear whether these tools provide a real measure of impact made up of IRIS and GIIRS, even though the industry aims to maximize positive social and environmental impacts, its tools do not in fact measure impact because the latter is too complicated and controversial to evaluate. To clarify, impact is defined as improvement to a person’s life or a social outcome *caused* by an intervention above and beyond what is attributable to other factors.¹²⁴ Thus impact assumes a causal relationship between program participation and change in the life of the participant. That is, what would have happened to a person who participated in a program (taking on a microfinance loan, for instance) if they had not done so?

Drawing a causal relationship between an intervention and a positive outcome is a fraught and imprecise process, however.¹²⁵ Consequently, the framework relies on *proxies* for impact, in the form of outputs: “Rigorous impact evaluation . . . is powerful, but onerous and expensive in practice. Many impact investors therefore

¹²² Ibid.

¹²³ Ibid.

¹²⁴ Syed Hashemi, “Beyond Good Intentions: Measuring the Social Performance of Microfinance Institutions,” *Focus Note 41*, CGAP (May 2007), 2.

¹²⁵ In answer to the question: “Why did we stop at outputs?” The *previous* IRIS FAQ’s answer: “The taxonomy captures only outputs, or direct, measurable performance indicators as this reduces the potential issues/concerns associated with the interpretation of those metrics” (<<http://iris-standards.org/faq>>). This website has been replaced, however, and the new FAQs no longer include this discussion.

settle for measuring ‘activities’ or ‘outputs’ (such as the number of bed-nets sold) rather than running control groups to measure the ‘impact’. Investors balance the need for rigorous impact evaluation against the need for simple, cost-effective ways of measuring this impact.”¹²⁶

As an example, rather than seeking out information relating to the substance or quality of stakeholder feedback, GIIRS would ask only whether such feedback is sought and through what mechanisms (e.g., annual stakeholder meetings, anonymous surveys). Or a company would be asked whether its product or service has a direct or indirect (or no) impact on economic equality or on health, for instance, but they would not be asked to substantiate that the impacts were in fact generated. It may therefore be misleading to label this sort of indicator a measure of “impact,” and charges of ineffectiveness could be leveled against the tools accordingly.

Further, for legitimacy purposes, it is important to ask: “Who decides on ‘the line between a level of detail that is too onerous to collect and one that is too superficial to be useful’¹²⁷ in defining social impact measures? And are outputs in fact good proxies for impact? Whose views are sidelined in that determination and at what cost? Does measuring outputs without a causation analysis mean that certain output-determining factors are not taken into account when aggregating ratings, such as a drought or a political crisis, or an injection of supplementary funds from a development agency that is perhaps more attuned to local needs, and thus better equipped to generate outputs?”

In addition, there are some aspects of performance that IRIS and GIIRS have only limited ability to measure or assign due value to, such as the social cohesion and political empowerment outcomes of an investment, which could, if included, affect a GIIRS rating. This issue will be explored in more detail below.

Indicators and knowledge production

The previous sections elucidated the roles that GIIRS and IRIS play in governance, as interventions that create structures for and influence decision-making by impact investors, investees, and official policy-makers, particularly during the market-building stage. These manifestations of indicator-power are relatively concrete. At a more abstract level, indicators also act as powerful knowledge producers, influencing not only decision outcomes, but also the very process of decision-making, fundamentally altering the values and the logics involved.

In the blended value context, GIIRS and IRIS behave like vessels that transport norms from philanthropy to commerce, and back. As these “norm-vessels” travel between problem-solving spaces, they shape and influence the ways in which problems are identified, framed, and addressed. In so doing, they become forces of knowledge production, story-tellers in their own right, which is quite a conceptual promotion for a (seemingly simple) numerical or star-based score.

¹²⁶ Emerging Asset Class (n. 1), 72.

¹²⁷ Emerging Asset Class (n. 1), 75.

Considering GIIRS and IRIS within the context of a larger and growing field of indicators, this section argues that indicators can change decision-making processes by altering the forms of knowledge that are relied upon by decision-makers. In this case, the effect is to introduce social logic to market actors and market logic to philanthropic actors. By drawing attention to the fact that efforts to socialize businesses and harness markets to solve global problems are not taking place in a vacuum, but rather alongside a move to marketize both philanthropy and international development, this section suggests that GIIRS and IRIS are setting sail across contested waters.

Altering forms of knowledge: Dissolving the dichotomy between social and commercial

Underlying the impetus to establish an impact investing industry is the view that capital should be unshackled from the conventional corporate forms and structures that have traditionally determined its allocation. Indeed, there is a resounding call among industry proponents to remove “the established systems of intermediating capital that force us into a binary choice between using for-profit investment purely to maximize financial return and using charity and subsidy when we seek to solve social problems.”¹²⁸ Standing firm in the belief that commercial capital can achieve at least as much as grants or subsidized funds to generate social and environmental benefits, industry advocates seek to do away with the notion that for-profit activities and social interests should be kept separate.¹²⁹ Viewing this separation as “a systematic barrier to the development of the impact investing industry,” they are critical of the policy and regulatory frameworks that sustain it, and seek, at a minimum, the “recognition that it is both economically feasible and morally legitimate to address social problems with for-profit investment.”¹³⁰ For impact investing proponents, such a shift is not only desirable but also inevitable: “The old binary system—the widely held belief that for-profit investment could only maximize financial return while social purpose could only be pursued through charity—is breaking down.”¹³¹

By giving stakeholders means by which to identify, measure, and compare social and environmental impact, indicator-based systems such as GIIRS and IRIS (and the structures established to disseminate them) convert social and environmental values from the philanthropic arena into measurable returns. In this way, they facilitate the production and recognition of a different kind of value, one that can be rated and transacted against in the commercial arena. In this way too, they participate in the “breaking down” of the barrier that separates the commercial

¹²⁸ Antony Bugg-Levine, Interview, @lliance for Philanthropy and Social Investment Worldwide, May 1, 2010, available at <<http://www.alliancemagazine.org/node/3283>>.

¹²⁹ Ibid.

¹³⁰ Ibid.

¹³¹ See <<http://www.globalimpactinvestingnetwork.org/cgi-bin/iowa/resources/research/72.html>>.

sphere from the social sphere. Conversely, indicators play an important role as vessels for communicating market logic and norms into the philanthropic context. Indeed, the adoption of performance indicators in the philanthropic sphere enables the “marketization of philanthropy,” or the use of market logic and mechanisms in developing strategies for the allocation of limited philanthropic funds.

Practically speaking, indicators are performance assessment aids used in making decisions that should result in more “effective” allocations and re-allocations of funds toward “better” performing projects (however defined/measured).¹³² By relying on the information indicators produce, donors expect to make more “impactful” allocation decisions.¹³³ The same should also hold true for commercial investors, as they become increasingly better equipped to assign transaction value to social and environmental impact.

More abstractly, indicators, as transporters of market logic, are deployed to tame the “emotionality” of donor decision-making. They serve a sobering function thanks to their perceived objective, reason, and/or science-based features.¹³⁴ Indeed, the market, and more specifically, market-based approaches to problem-solving, tend to be viewed as having quasi-scientific authority, and the “discipline” associated with “hard-nosed business management approaches” is seen as a “useful antidote to the often emotional appeals”¹³⁵ that can accompany philanthropic funding decisions.¹³⁶

It is worth highlighting that indicators are not the only device being used to marketize philanthropy;¹³⁷ in fact, this effort has been undertaken on a number of

¹³² See the chart entitled “Nonprofit Marketplace Information Network” in Marie O’Flanagan, Jacob Harold, and Paul Brest, “The Nonprofit Marketplace: Bridging the Information Gap in Philanthropy,” William and Flora Hewlett Foundation, 2010 (hereinafter Nonprofit Marketplace), available at <<http://www.hewlett.org/uploads/files/whitepaper.pdf>>, 11.

¹³³ See, e.g., Melinda T. Tuan, “Measuring and/or Estimating Social Value Creation: Insights into Eight Integrated Cost Approaches (Bill & Melinda Gates Foundation—Impact Planning and Improvement) (December 15, 2008) (hereinafter Tuan), 25, and Nonprofit Marketplace” (n. 132), 26.

¹³⁴ Wendy Nelson Espeland and Mitchell L. Stevens, “A Sociology of Quantification,” *European Journal of Sociology* XLIX 3 (2008): 401–36, explaining that “raw” information is typically collected and compiled by low-level decision-makers and that as it moves up the organizational hierarchy, it is transformed and simplified in ways that remove assumption, discretion, and ambiguity, resulting in “uncertainty absorption” that makes information appear more robust than it actually is, infusing it with more authority than is warranted.

¹³⁵ Tuan (n. 133), 4, citing Lynn A. Karoly, M. Rebecca Killburn, James H. Begelow, Jonathan P. Caulkins, and Jill S. Cannon, “Assessing Costs and Benefits of Early Childhood Intervention Programs,” *RAND Monograph* (2001), 5.

¹³⁶ Tuan (n. 133), 5; see also Katherine Fulton, Gabriel Kasper, and Barbara Kibbe, Monitor Institute, “What’s Next for Philanthropy? Acting Bigger and Adapting better in a Networked World,” July 2010, 4: “In many cases, whether a grant or gift is ‘effective’ may matter less than the values it represents to the donor, the personal commitments it reflects, or the web of relationships it helps to maintain. As scholar Peter Frumkin observed, ‘At its core, [philanthropy] is about expressing values, not outcomes.’”

¹³⁷ For other examples, see Social Returns on Investment (SROI) methodologies used by non-profits that assign a dollar value to the social benefits earned for each dollar donated (e.g., \$1 invested = \$3 social return). See also, Lucy Heady, “Social Return on Investment” Position Paper of New Philanthropy Capital, April 2010, available at <http://www.philanthropycapital.org/publications/improving_the_sector/charity_analysis/sroi_position_paper.aspx>; see also the tool developed by the Hewlett Foundation, which includes separate star ratings for stakeholder feedback (experts, volunteers,

fronts, including linguistically. As an example, a report by the William and Flora Hewlett Foundation and McKinsey illustrates how market systems can generally serve as a blue print for philanthropy; here, donors are re-labeled as “social investors” who participate in a “nonprofit marketplace” where they can be (more or less) effectively connected with recipients.¹³⁸ The authors explain:

“Recognizing the parallels with the capital markets in terms of information and investment flows, many observers have described philanthropy as the investment capital of the ‘non-profit marketplace.’ In this marketplace, donors (the equivalent of investors) connect with nonprofit organizations (suppliers) to serve beneficiaries (end consumers), often with assistance from a variety of market intermediaries that provide information, transaction support, advice, education, and networks.”¹³⁹ In the report’s concluding recommendations, the authors “urge donors to adopt more of an investment mindset. This can start with a simple shift in language: from ‘giving to charity’ to ‘investing in high-performing nonprofits.’¹⁴⁰ But a true investment mindset goes beyond semantics: it implies a move toward active engagement and high expectations.”^{140a}

Formulating and structuring expectations concerning returns on charity-as-social-investing is an important feature of the philanthropy-marketization trend. And this is precisely the dimension of norm-transfer that tools such as GIIRS and IRIS are intended to support, given that they formulate and disseminate a returns-based approach for social investing. These tools actively participate in the norm-exchange (and merger?) between commerce and philanthropy that is encouraged by industry proponents, as explained by Maximilian Martin (founder of Impact Economy, a social finance and strategy firm):

“One of the key changes in the Impact Economy consists of the way people allocate investment capital... There are two dimensions to this: in philanthropy, capital will be deployed increasingly based on the principle of *creating social value*, as opposed to nurturing donor-grantee relationships. In business, capital will increasingly be deployed *based on values* instead of market anonymity... This ‘value vector’ drives the transformation of both business and philanthropy.”¹⁴¹

As they shuttle between the problem-solving spaces of commerce and philanthropy, indicators behave like vessels that are charged with transporting different and possibly conflicting values and norms. The shuttling also creates *new* forms of value and knowledge; it reshapes decision-making processes and the way that problems are identified, selected, conceptualized, and eventually addressed: “[I]ndicators have the potential to prompt changes in beliefs about the world,

peer organizations, and beneficiaries), Nonprofit Marketplace (n. 132), 45; *see also* Tim Ogden, “The Worst (and best) Way to Pick a Charity This Year,” Press Release issued December 1, 2009, describing a new ranking by Charity Navigator that emphasizes charity effectiveness and transparency.

¹³⁸ Nonprofit Marketplace (n. 132), 7.

¹³⁹ Ibid.

¹⁴⁰ Ibid., 52

^{140a} Ibid., 52

¹⁴¹ Maximilian Martin, “The Impact Economy: Beyond Profit and Philanthropy” (forthcoming and available on SSRN), 5.

including beliefs about the moral implications of various social outcomes, about relationships between actions and outcomes, and about the beliefs or actions of other actors.”¹⁴² Although indicators are not the only device in use for bridging the social/commercial divide, because of their belief-altering and knowledge-producing attributes, they may be among the more powerful.

Challenges to the knowledge-building function of indicators

Contesting marketization

As loud as it may be, the call to dissolve the social/commercial binary and to create structures that unleash private capital for social purposes does not have universal resonance. Critics of the blended value proposition coming from the non-profit and international development spheres, as well as from academia, are concerned about the desire to break down the border between profit and social purpose, and also about the adoption of market-driven solutions for achieving social and environmental goals. In contesting the spread of market logic into the social sphere, author Michael Edwards¹⁴³ queries whether in the future, there will be “anyone left who responds to something other than a market incentive” and worries that the move toward greater reliance on market models and tools will “divert attention away from the deeper changes that are required to transform society, screen out difficult but vital questions for research and policy-making, reduce decision-making to an inappropriate bottom line, and lead us to ignore the costs and trade-offs involved in extending market mechanisms into the social world.”¹⁴⁴

Of further concern to Edwards is the effect of marketization on conceptions of “social value”; in particular, “[t]he individualism of the marketplace is not the same as collective action and mutuality. These are radically different philosophies of action. Customers, clients, or consumers do not have the same rights and responsibilities as citizens. Having more choice between suppliers is not the same as economic and political empowerment.”¹⁴⁵ Their function as quantifiers and commodifiers of social value does little to ingratiate Edwards to indicators. In fact, he distinguishes “metrics-seekers” from “social justice-seekers”; the former are participants in a “machine-based system” where the production and demand for data-driven indicators is based on the primacy of standardization, without which the

¹⁴² See Kevin Davis and Benedict Kingsbury, “Indicators as Interventions: Pitfalls and Prospects in Supporting Development Initiatives,” Rockefeller Foundation, December 2011 (hereinafter *Indicators as Interventions*) available at <<http://www.rockefellerfoundation.org/news/publications/indicators-interventions>>, 17.

¹⁴³ Michael Edwards was previously the Director of the Governance and Civil Society Unit at the Ford Foundation and a World Bank staff member. He is the author of *Small Change: Why Business Won't Save the World* (San Francisco: Berrett Koehler, 2010).

¹⁴⁴ Michael Edwards, “Oil and Water or the Perfect Margarita? Where is the ‘Social’ in the ‘Social Economy?’” *The Philanthropist* 22:2 (2009) (hereinafter *Oil and Water*): 77.

¹⁴⁵ *Savior or Emperor* (n. 14), 10.

system would fail.¹⁴⁶ The latter, he argues operate on the correct premise that, “the world of social change is an ecosystem, not a machine; it is characterized by diversity, emergence, unforeseen events, organic growth, and evolution, rather than ideas of predictability.”¹⁴⁷

Similar concerns are echoed in the international development context. Former USAID Administrator, Andrew Natsios, identifies a “counter-bureaucracy” made up of “U.S. government agencies charged with command and control of the federal bureaucracy through a set of budgeting, oversight, accountability, and measurement systems that have grown over several decades to a massive degree.”¹⁴⁸ He diagnoses this counter-bureaucracy with a “very bad case of Obsessive Measurement Disorder,” meaning “an intellectual dysfunction rooted in the notion that counting everything in government programs (or private industry or increasingly some foundations) will produce better policy choices and improved management.”¹⁴⁹ Natsios associates this rather frightening-sounding condition with a tendency to rely on business management tools in pursuing development goals.¹⁵⁰ He is highly critical of the shift in decision-making and evaluation toward measurable outputs, in part because “this means more funding for those development sectors in the hard sciences, such as public health and medicine, and less to the soft sciences such as democracy and governance programs.”¹⁵¹ He argues that though the latter are “less visible, harder to measure, and much slower to demonstrate success than the hard sciences,” the “central principle of development theory” is that “those development programs that are most precisely and easily measured are the least transformational, and those programs that are most transformational are the least measurable.”¹⁵² He asserts that “[e]ssentially, measurability should not be confused with development significance.”¹⁵³

Professor Ananya Roy is concerned with the “financialization of development.”¹⁵⁴ This term adds important layers to the marketization phenomenon,

¹⁴⁶ Michael Edwards, Phone Interview, August 6, 2010. Notes on file with author. On the risk of seeking to de-politicize inherently political processes, see Katharina Pistor, “The Standardization of Law and its Effects on Developing Economies,” *American Journal of Comparative Law* (Winter 2002), and Frank Upham, “Mythmaking in the Rule-of-Law Orthodoxy,” in Thomas Carothers (ed.), *Promoting the Rule of Law Abroad: In Search of Knowledge* (Washington DC: Carnegie Endowment for International Peace, 2006).

¹⁴⁷ Michael Edwards, Phone Interview, Friday, August 6, 2010 (notes on file with author). Note that this distinction points to one reason why the notion of impact (as the change caused by an intervention) is more useful than outputs, as it is broader and more capable of assigning importance to the institutional, economic, and cultural settings in which programs are implemented.

¹⁴⁸ Andrew Natsios, “The Clash of the Counter-bureaucracy and Development,” Center for Global Development, July 2010, available at <<http://www.cgdev.org/content/publications/detail/1424271>> (hereinafter Natsios), 4.

¹⁴⁹ Natsios (n. 148), 4.

¹⁵⁰ Natsios (n. 148), 4 and 6.

¹⁵¹ Natsios (n. 148), 10.

¹⁵² Natsios (n. 148), 4 and 63.

¹⁵³ Natsios (n. 148), 13.

¹⁵⁴ Ananya Roy, *Poverty Capital: Microfinance and the Making of Development*, (New York: Routledge, 2010), 31 (hereinafter *Poverty Capital*). Roy outlines three paradigms of microfinance: “Yunus’s emphasis on a rights-based, pro-poor approach to microfinance; ‘creative capitalism’ strategies that position the poor as a lucrative market; and sharp critiques of microfinance that reject its impacts

and can easily encompass impact investing; it describes the process by which microfinance, through its repackaging as “financial services for the poor,” has evolved into “a new global industry that can be integrated into financial markets”; it establishes a direct link between “the dispersed millions of impoverished micro-entrepreneurs in the global South and centralized nodes of financial power in the global North.”¹⁵⁵ Roy defines “poverty capital” as a “subprime frontier where development capital and finance capital merge and collaborate such that new subjects of development are identified and new territories of investment are opened up and consolidated.”¹⁵⁶ She asks, “[w]ill the poor benefit from such integration or will their inclusion take place in highly exploitative or predatory ways?”¹⁵⁷

Roy draws attention to the role of indicators as both inputs and outputs of the financialization of development: “Poverty capital is not only the practice of lending and producing wealth. It is also the practice of producing knowledge . . . It is here that a metrics of risk assessment and management is forged; it is here that the poor are classified and categorized; and it is here that more generally the business of poverty comes to be ‘financialized,’ or transformed into a set of financial benchmarks and indicators.”¹⁵⁸

By organizing and producing knowledge about how to define, frame, and solve social problems, indicators acquire the power to empower and disempower, to favor and to stigmatize, which is a form of governance, one characterized by a fundamental experiential mismatch; simply, the people affected by the issue are distinct from those who formulate the issue.¹⁵⁹ Upon entering this governance space, indicators are inserted into a tension-ridden “geography of development ideas,”¹⁶⁰ in which the global North produces norms and framings that are then transplanted and imposed on the global South. Here, indicators become “the currency of poverty experts, those who are authorized to produce authoritative knowledge about poverty and its alleviation; they eventually come to be understood as certain, legitimate, and undeniably correct . . . They are taken up by those who implement development and by those upon whom development is conferred.”¹⁶¹

Contesting measurement: Stars or stories?

In thinking about alternatives to metrics, consider the call by Natsios to exempt certain programs (e.g., institution building, certain service delivery, economic growth, technical assistance, scholarship, research, and conflict mitigation programs)

on poverty. The three paradigms are about much more than microfinance. They represent the struggles within development over issues such as social interests, the regulation of markets, and the nature of capitalism.” At 23.

¹⁵⁵ Poverty Capital (n. 154), 30–1.

¹⁵⁶ Ibid.

¹⁵⁷ Poverty Capital (n. 154), 32 and 195.

¹⁵⁸ Poverty Capital (n. 154), 31.

¹⁵⁹ Indicators as Interventions (n. 142), 1.

¹⁶⁰ Poverty Capital (n. 154), 196.

¹⁶¹ Poverty Capital (n. 154), 31–2, citing Imram Martin and Shamim Ara Begum, “Asset-ing the Extreme Poor: Experiences and Lessons from BRAC Project,” RED BRAC Economic Studies, XVIII. See also Governance by Indicators (n. 3), explaining that while indicators may at first be subject to great scrutiny and debate, a year or two after their release, “the indicator is presented simply as a fact that describes the country’s situation, with virtually no discussion about the source of the data or the nature of the indicator itself.”

from quantitative measurement.¹⁶² He says that these programs, amongst others, “are suffering from the misapplication of short-term indicators,”¹⁶³ and recommends that an alternative, story-based model be adopted: “Field impact evaluations—where a team of evaluators conducts extensive interviews with aid beneficiaries across several countries to determine whether a given program approach is producing the intended results—should replace” the evaluation mechanisms currently in use, including indicators.¹⁶⁴

As the development and support of GIIRS and IRIS demonstrates, however, stories often lack the credibility of market tools in the eyes of (Northern) capital custodians. This is well illustrated in Roy’s telling of how communicating value through stories, rather than through stars or a score, loses the fight for “truth” and knowledge control. From an interview with a Bangladeshi economist at the Consultative Group for Assistance to the Poor (CGAP) (a think tank housed within the World Bank that focuses on developing knowledge around and standardizing the microfinance sector¹⁶⁵): “Microfinance is a good example of knowledge appropriation . . . How did it go from Bangladesh to Boulder? We created microfinance and then lost it.” And later, “[w]e are good at creating snapshots and telling stories. It is like poetry. But it does not come across as technical. It seems to lack economic rationality. It lacks generalization. It cannot become universal. And so we lost it to those who can make this poetry both technical and universal.”¹⁶⁶

Roy also relays her conversation with Imran Martin, Director of BRAC’s Research and Evaluation Division from 2005. BRAC, one of the largest NGOs in the world, is based in Bangladesh. It provides microfinance services but embeds them in a vast array of social protection and human development services, thus adopting a holistic approach to development.¹⁶⁷ Martin describes CGAP as “global and very influential, especially through its indicators and monitoring tools . . . Even our Grameen Trust partners often have to follow CGAP indicators in order to procure funding from donors. How do we create an alternative? . . . We have demonstrated success with these ideas in Bangladesh and yet in global forums no one wants to believe it can be done. It’s frustrating.”¹⁶⁸

Through their apparent objectivity and universality, indicators validate, absorb, and trump indigenous truth-making and story-telling, as Grameen Bank founder Muhammad Yunus explains, “[s]o the process works like this: that our indigenous ideas have to be validated by the West before we can accept them. We don’t know

¹⁶² Natsios (n. 148), 66.

¹⁶³ Ibid.

¹⁶⁴ Ibid., 67. One approach would be to adopt the Most Significant Change (MSC) program evaluation model, perhaps the most radical alternative to an indicator. MSC involves the collection of “significant change” stories from field level staff and stakeholders, and the selection of particular stories by other actors within the organizational hierarchy. The stories are then read aloud and discussed in groups. See “What is MSC,” Monitoring and Evaluation News, available at <<http://mande.co.uk/special-issues/most-significant-change-msc>>.

¹⁶⁵ <<http://www.cgap.org/p/site/c/aboutus>>.

¹⁶⁶ Poverty Capital (n. 154), 197.

¹⁶⁷ Poverty Capital (n. 154), 25.

¹⁶⁸ Poverty Capital (n. 154), 123.

how to sell our own ideas.”¹⁶⁹ These quotes underscore an important feature of indicators, namely that as their levels of acceptance and use increase, so too do the depth and texture of their governance role. Indicators can assume policing functions, as gate-keepers for development knowledge, allowing some forms of knowledge in and keeping others out. This authority is granted even though indicators might not accurately represent the change-oriented activities they purport to measure or the interests of those who are supposed to benefit from such activities. Where funding is tied to indicator performance, funding-seekers might pander to indicators, designing their activities to receive positive indicator-based assessments, rather than positive assessments based on their *own* beliefs as to what constitutes success.¹⁷⁰ The interviews strongly suggest that such indicator-evolutions can prompt contestation; whether such contestation is effective is less clear, however.

As indicators shuttle between philanthropy and commerce, it is important to ask whether the social norms they transport and leave behind are thinned following each “passage.” Is there a risk that the quantification and asset-ization of social value—into impact—will be accompanied by only limited transformation on the commercial side—in terms of achieving meaningful change in the way business does business? While these passages may not seem risky at first glance, some critics certainly worry that one set of norms will come to dominate. One troubling outcome of market-side domination is that we may become less critical of market failures, and possibly even less able to identify them. Given that the market has historically, and certainly of late, been less than reliable when it comes to delivering social and environmental goods, it may be wise to focus on keeping our critical skills sharp.

Philanthropic and market approaches each produce different identifications of problems, different solutions for those problems, and different assessments of success or failure. If the two approaches become merged, as a corollary of the dissolution of the social/commercial dichotomy, what nuances and innovative spaces might be lost? Whose voices could go unheard? Edwards worries that pursuing social purposes through market models will hurt philanthropy and the causes it supports:

By mixing these different rationalities together, are we producing the ‘perfect margarita’ of ‘creative capitalism’ or pouring ‘oil on water’ in ways that pollute civil society and reduce its transformative potential? That’s the question, and unless we are clear on the answer, we may find ourselves facilitating the rise of the economic society rather than the social economy. The consequences of that would, I think, be profound, in terms of a society driven by finance and defined by economics instead of an economy transformed—or at least held accountable—by social or civil society values and priorities.¹⁷¹

¹⁶⁹ Cited in Poverty Capital (n. 154), 125. This statement was made at the closing session of the forty-fifth Grameen Dialogue, held in 2003.

¹⁷⁰ For further support for the claim that indicators can distort local policymaking and actions, see Indicators as Interventions (n. 142), 34.

¹⁷¹ Oil and Water (n. 144), 76.

These lines of inquiry highlight that impact indicators, as knowledge producers and norm-vessels, do not operate in a vacuum. The problem-solving spaces that they travel between—philanthropy, development, and commerce—are being fought over and redefined by a broad range of stakeholders and truth-makers. Thus GIIRS and IRIS will travel across waters that are not only charted, but also highly contested. And, given the position of industry leaders on dissolving the dichotomy between the social and the commercial, it is unlikely that these tools have set sail as neutral conflict mediators.

Conclusion

As impact investing establishes itself as an industry, indicators have a vital governance role to play. By organizing and systematizing impact investments, GIIRS and IRIS both structure and influence decision-making by investors and investees. In so doing, they build a market for impact investing, making possible its graduation from idea to industry to asset class. GIIRS and IRIS also play an important role as interventions that can influence, stimulate, and even incubate regulation by governmental actors, thus contributing to the official policy landscape into which impact investing could be inserted.

Both of these governance roles give rise to contestation. As market builders, IRIS and GIIRS may face infrastructural challenges to their efficacy. They may be discredited for being premature if the capital absorption capacity of mission-driven businesses is too low to generate meaningful returns (financial or social). Infrastructural limitations may also impair the ability of smaller businesses to secure high ratings, which creates another avenue for contestation. Further, though the IRIS/GIIRS impact assessment framework does not currently have direct competition within the indicator-sphere, other more established performance assessment mechanisms might stake a claim as better performers of the governance functions currently exercised by IRIS and GIIRS.

In terms of regulatory decision-making, the IRIS/GIIRS impact assessment framework may face structural challenges to its legitimacy. Contestations could emerge concerning their US-centric and private entity-dominated governance structures or the low level of engagement with beneficiaries. Additional challenges could arise if stakeholders clash because of conflicting interests, or because they hold differing views as to what constitutes “impact” or as to acceptable levels of tradeoff between social, environmental, and financial returns. Yet another challenge pertains to the misleading-ness of advertising IRIS and GIIRS as capturing “impact” given the tools’ (acknowledged) inability to measure impact, as opposed to outputs.

While these infrastructural and structural challenges may be daunting, they could potentially be overcome through re-design, particularly since IRIS and GIIRS are still being developed and fine-tuned. However, in the process of building a blended value world, a third governance function of indicators (including, but not limited to GIIRS and IRIS) creates challenges that re-design alone would likely be unable to address. As this chapter argues, indicators serve as norm-vessels that

transport social and environmental “value” into commercial investing, and market logic into philanthropy. As they travel between problem-solving spheres, indicators bridge and dissolve the divide between the social and the commercial. In so doing, they influence the very process of decision-making, creating new forms of knowledge that structure the identification of stakeholders and beneficiaries as well as the formulation of social and environmental problems and viable (investable) solutions.

This governance role is more normatively problematic than the market-building or regulation-shaping functions because it shines light on the subjective nature of indicators and how they can be used to shift paradigms to create new “truths.” Bridging the space between commerce and philanthropy entails more than harnessing previously untapped pools of capital to solve social and environmental problems. It can also entail converting “the microcapital of the poor into global financial flows”¹⁷² or identifying and commodifying value from within poverty. Additionally, it can mean introducing a structure of expected returns into philanthropy, a system that may achieve more by developing problem-solving activities without expectation of short-term or measurable outcomes, outputs, or impacts. Finally, it creates the risk of further entrenching a monopoly by the global North over poverty-knowledge, which is already increasingly expressed through bottom lines, whether single or multiple. Each of these dimensions of the bridging process gives rise to contestation.

Impact investing industry advocates package this last set of challenges as a “reputational risk,” adding that “mission drift . . . and exploitation are legitimate concerns” and recommending that impact investors “ensure [that] the right metrics are in place to monitor their portfolio companies.”¹⁷³ Relying on the “right metric” is a tautological solution, however. While it may not be possible to address at this stage, this challenge nevertheless needs to be well understood, and that involves recognizing that IRIS and GIIRS are not neutral, nor are they setting sail on calm waters. As well-intentioned, useful, consensus-driven, and exciting as they may be, they inevitably participate in capital allocation and knowledge production processes that are deeply contested geographically, economically, and morally.

¹⁷² Poverty Capital (n. 153), 32.

¹⁷³ Emerging Asset Class (n. 1), 68.

PART V

REGULATING INDICATORS

This page intentionally left blank

Accountability in the Generation of Governance Indicators

*Nikhil K. Dutta**

Introduction

Decision-making is difficult in a complicated world. As the number of considerations involved in making a decision increases, it becomes harder to gather all the information needed to make an informed choice. Proliferating considerations also introduce uncertainty about how to weigh different factors in making a final choice, and complicate the process of justifying decisions to others, especially when some of the considerations involved in the decision appear inherently subjective.

Measures that summarize complex realities can make decision-making easier. Scholars refer generally to such measures as “indicators.”¹ If indicators simplify decisions, though, they often do so by transferring the “heavy lifting”—in terms of gathering information, exercising judgment, and justifying choices made—from the point of decision to the point at which indicators are generated. In international settings, indicators are becoming more popular, from the United Nations (UN) Human Development Index² to Moody’s and Standard & Poor’s sovereign credit ratings³ to human rights indicators generated by a variety of organizations.⁴ Indicators are vulnerable to a variety of criticisms: that they are “by definition a

* Some material in this chapter first appeared in 22 *Fla. J. Int’l L.* 401 (2010). I would like to acknowledge with gratitude the advice and encouragement of Christopher Bradley, Kevin E. Davis, Angelina Fisher, Benedict Kingsbury, and Richard Stewart. All errors that remain are my own.

¹ Willem F.M. de Vries, “Meaningful Measures: Indicators on Progress, Progress on Indicators,” *International Statistical Review* 69 (2001): 313, 315 (defining “statistical indicators as single numbers . . . that try to capture a more or less complex reality”).

² UN Development Programme (UNDP), Human Development Reports, available at <<http://hdr.undp.org/en/statistics/>> (accessed August 5, 2011).

³ Richard Cantor and Frank Packer, “Sovereign Credit Ratings,” *Current Issues in Economics and Finance* (June 1995): 1. (“In recent years, the demand for sovereign credit ratings—the risk assessments assigned by the credit rating agencies to the obligations of central governments—has increased dramatically.”)

⁴ See Rajeev Malhotra and Nicolas Fasel, “Quantitative Human Rights Indicators—A Survey of Major Initiatives” (March 2005) (discussing major attempts to develop quantitative human rights indicators), available at <<http://www.aboakademi.fi/instut/imr/research/seminars/indicators/Background.doc>> (accessed August 5, 2011).

simplification of reality”;⁵ that they represent a “widespread turn to ‘accounting culture’ in which tests of measurability often prevail over accurate and contextually sensitive assessments of substance or actions”;⁶ that they rely on information that “contain[s] biases that . . . can distort ratings.”⁷ When actors base important decisions on indicators, moreover, concerns about the “accountability deficit in the growing exercise of transnational regulatory power”⁸ suddenly apply to indicator generation, as well. This chapter aims to explore the determinants of accountability in the generation of governance indicators.

For the purposes of this chapter, a “governance indicator” is any indicator that measures an aspect of good governance. Such indicators (1) select some aspect of the relationship between a government and its society; (2) identify a normative framework within which this relationship can be judged; and (3) measure how well the actual relationship within a country adheres to the normative framework. Governance indicators thus do not explore inter-governmental relations like trade negotiations, extradition practices, or representation at international bodies, or study purely private practices like family life, religious organization, production, or trade—except to the extent that these phenomena reveal something about the way governments relate to their peoples.⁹

The chapter postulates two processes that might determine levels of accountability in the generation of governance indicators. The first is the Demand Hypothesis: that the response of generators to demands for accuracy or influence from users and targets of indicators determines observed levels of accountability. The second is the Supply Hypothesis, which predicts that indicator generators provide higher levels of accountability in order to attract the patronage of users. The chapter tests how well each of these hypotheses explains observed accountability in the generation of three indicators: the International Country Risk Guide ratings produced by the PRS Group, a for-profit private company;¹⁰ the Freedom in the World ratings generated

⁵ De Vries, “Meaningful Measures,” 315.

⁶ AnnJanette Rosga and Margaret L. Satterthwaite, “The Trust in Indicators: Measuring Human Rights,” *Berkeley Journal of International Law* 27 (2009): 253, 256.

⁷ Kenneth A. Bollen, “Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984,” in *Human Rights and Statistics* (Thomas B. Jabine and Richard P. Claude, eds, University of Pennsylvania Press, 1992), 188, 201.

⁸ Benedict Kingsbury et al., “The Emergence of Global Administrative Law,” *Law & Contemporary Problems* 68 (Summer/Autumn 2005): 15, 16.

⁹ Consequently, indicators that measure development levels, like the UN Human Development Index (HDI), should not really be considered governance indicators, though government policies certainly affect their levels. See UN Development Programme, Human Development Reports; see also de Vries, “Meaningful Measures,” 322–3 (describing method by which HDI is calculated). Indicators that measure economic or financial performance within a country, such as the International Country Risk Guide Economic Risk Index, are similarly excluded from our remit. See The PRS Group, International Country Risk Guide Methodology 7 (2011) (“The overall aim of the Economic Risk Rating is to provide a means of assessing a country’s current economic strengths and weaknesses”), available at <<http://www.prsgroup.com/PDFS/icrgmethodology.pdf>> (accessed August 5, 2011) (hereinafter ICRG Methodology Manual).

¹⁰ The PRS Group, *International Country Risk Guides (ICRG)*, <<http://www.prsgroup.com/ICRG.aspx>> (accessed August 5, 2011) (hereinafter ICRG Overview).

by Freedom House, a non-profit non-governmental organization;¹¹ and the Minorities at Risk variables developed by the MAR Project, an academic Project at the University of Maryland.¹²

The Demand Hypothesis of accountability

An administrative body can provide accountability by “meet[ing] adequate standards of transparency, participation, reasoned decision, and legality, and by providing effective review of the rules and decisions they make.”¹³ The Demand Hypothesis predicts that as decisions based on indicators involve progressively higher stakes, the intensity of demands by users for accuracy¹⁴ (and possibly accountability), and by targets for accountability, should increase. All else being equal, this should result in greater levels of accountability provision by indicator generators.

The Demand Hypothesis makes sense because the main types of actors who might be moved by self-interest to demand enhanced accountability are the users and targets of indicators.¹⁵ When decisions based on indicators affect targets, they should seek greater influence over how indicator levels are set—either to secure improved accuracy, or just to inflate or deflate their scores. Consequently, targets should demand increased levels of accountability in indicator generation when the stakes of decisions based on indicators rise. As for users, insofar as they want to base their decisions on accurate information and increased accountability seems likely to improve accuracy, one would expect them to demand enhanced levels of accountability, as well.

This latter link, between accountability and accuracy, seems a plausible one. After all, each of these components of accountability mentioned above should, if bolstered in indicator generation, enhance confidence in indicators’ accuracy. Participation—or considering the views and information of affected parties—allows those interested in the accuracy of indicators to share with indicator generators methodologies and data that they might otherwise neglect. Such participation can be enhanced in effectiveness if those who participate have full access to

¹¹ FreedomHouse.org, Freedom in the World, <<http://www.freedomhouse.org/reports>> (accessed August 5, 2011) (hereinafter Freedom in the World).

¹² MAR, *The MAR Project*, <<http://www.cidcm.umd.edu/mar/>> (accessed August 5, 2011).

¹³ Kingsbury et al., “The Emergence of Global Administrative Law” (n. 8), 17.

¹⁴ Throughout this chapter, we will refer to accuracy in governance indicators as something indicator users or indicator targets seek. Such references should not be taken as presumptions that governance can accurately be measured or that accuracy is an inherent quality of indicators. Rather, this chapter is predicated only on the assumption that indicator users, generators, and targets will perceive indicators to be more or less accurate, and that such perceptions may motivate their behavior.

¹⁵ We will use “indicator users” to refer to the aid agencies, journalists, academics, investors, and others who use governance indicators to make decisions or substantiate claims. “Indicator generators” refers to the entities that produce governance indicators, whether they be aid agencies, non-governmental organizations, academics or universities, or private companies. “Indicator targets” refers to those whom governance indicators measure—that is, states, businesses, civil society organizations, and populations.

information about how indicators are set—that is, if indicator generators are transparent about their methodologies. Reason-giving—or explaining how indicator-generating processes are applied to information to produce an indicator’s particular values—allows outsiders to judge whether decisions are actually being made based on the processes and information revealed by greater transparency. The opportunity to appeal final decisions by an indicator generator as to the processes or information used to produce indicators, or the levels at which indicators are ultimately set, allows users and targets of indicators to circumvent institutional myopia, inertia, or indifference and instead plead their case before a disinterested arbiter. Finally, legality works in tandem with the other accountability components to provide assurance that individuals within a generating agency are actually following the processes revealed by transparency, carrying out the actions agreed upon through participation, or implementing changes suggested by an arbiter. In short, if users want to bolster the accuracy of governance indicators, improving accountability seems a good means to that end.

The three test cases

This section tests the Demand Hypothesis with reference to three sets of governance indicators: (1) the International Country Risk Guides developed by The PRS Group, a for-profit business; (2) the Freedom in the World rankings produced by Freedom House, a non-profit NGO; and (3) the Minorities at Risk indicators developed by the Center for International Development and Conflict Management at the University of Maryland. These indicators have been chosen to vary not only as to the nature of each indicator generator, but with respect to the predominant uses of each indicator.

International country risk guides, PRS Group

The PRS Group is a for-profit company that was founded in 1979.¹⁶ Its business largely revolves around two products: Political Risk Services (PRS) ratings, and International Country Risk Guides (ICRG).¹⁷ Both the PRS and the ICRGs are country risk ratings, but only the ICRGs explicitly measure the present state of governance in countries.¹⁸

¹⁶ The PRS Group, *About Us*, <http://www.prsgroup.com/AboutUs_Overview.aspx> (accessed August 5, 2011).

¹⁷ Ibid.

¹⁸ PRS ratings assess “[t]welve factors . . . from an 18-month forecast perspective,” and “[f]our additional factors . . . from a five-year forecast perspective,” establishing “likely changes in the level of political turmoil and 11 types of government intervention that affect the business climate”—implying that, strictly speaking, the PRS reports assess future, not present, governance. The PRS Group, *PRS Methodology*, <http://www.prsgroup.com/PRS_Methodology.aspx> (accessed August 5, 2011).

In 2010, the ICRGs included risk ratings for 150 “countries important to international business”;¹⁹ these ratings assess the degree of risk associated with each of 22 political, financial, or economic risk components before aggregating the components to provide a risk rating for each risk category as well as a composite risk rating.²⁰ Based on its aggregate political risk score, each country may be classified as “Very High Risk,” “High Risk,” “Moderate Risk,” “Low Risk,” or “Very Low Risk.”²¹

The ICRG ratings are purportedly “[u]sed by institutional investors, banks, multinational corporations, importers, exporters, foreign exchange traders, shipping concerns, and a multitude of others,”²² including “over 80 percent of the top companies in the world (as ranked by Fortune).”²³ The PRS Group also offers “sizable discounts on its business products . . . especially for academic use,”²⁴ and academics have used ICRG ratings to analyze the relationship between governance and a variety of dependent variables.²⁵

The ICRG ratings are also incorporated into other governance indices, the most prominent of which are the World Bank’s widely respected and influential Worldwide Governance Indicators (WGIs).²⁶ The WGIs comprise six governance indicators formed by aggregating several hundred individual variables drawn from a few dozen different sources.²⁷ Among its uses, the Millennium Challenge

¹⁹ ICRG Overview (n. 10), 10.

²⁰ The PRS Group, *ICRG Methodology*, <http://www.prsgroup.com/ICRG_Methodology.aspx> (accessed August 5, 2011).

²¹ ICRG Methodology Manual (n. 9), 7.

²² *Ibid.*, 1.

²³ The PRS Group, *FAQ*, <<http://www.prsgroup.com/FAQ.aspx>> (accessed August 5, 2011).

²⁴ The PRS Group, *Academic Titles*, <<http://www.prsgroup.com/academictitles.aspx>> (accessed August 5, 2011).

²⁵ For example, ICRG ratings have been used to analyze the impact on growth of secure property rights and contract enforcement: Christopher Clague et al., “Institutions and Economic Performance: Property Rights and Contract Enforcement,” in *Institutions and Economic Development: Growth and Governance in Less Developed and Post Socialist Countries* (Christopher Clague ed., The Johns Hopkins University Press, 1997), 67; the relationship between corruption and public investment: Vito Tanzi and Hamid Davoodi, “Corruption, Public Investment, and Growth,” in *The Welfare State, Public Investment and Growth* (Hiroyuki Shibata & Toshihiro Ithori eds, Springer, 1998), 41; and the effects of aid flows on the quality of governance: Stephen Knack, “Aid Dependence and the Quality of Governance: Cross-Country Empirical Tests,” *Southern Economic Journal* 68 (2001): 310.

²⁶ See, e.g., Steven Radelet, *Challenging Foreign Aid: A Policymaker’s Guide to the Millennium Challenge Account* (Center for Global Development, 2003), 34 (characterizing the WGIs as “the most comprehensive and best quality database available on governance indicators”); Christine Arndt and Charles Oman, *Uses and Abuses of Governance Indicators* (OECD Development Centre, 2006), 49, (describing the WGIs as “probably the most carefully constructed and widely used indicators”).

²⁷ Kaufmann et al., “Governance Matters VIII: Aggregate and Individual Governance Indicators 1996–2008,” (World Bank, Policy Research Working Paper No. 4978 June 2009), 7, available at <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1424591> (accessed August 5, 2011) (“For the 2008 round of the WGI, we rely on a total of 441 individual variables measuring different dimensions of governance. These are taken from 35 different sources, produced by 33 different organizations.”) The WGIs measure (1) voice and accountability; (2) political stability and absence of violence; (3) government effectiveness; (4) regulatory quality; (5) rule of law; and (6) control of corruption. *Ibid.*, 6.

Corporation (MCC)²⁸ employs five WIGs as criteria to select countries that will receive development grants.²⁹ At least one ICRG political risk component plays a role in every WGI and 10 of the 12 components are included overall.³⁰ Five of the MCC's 17 criteria are derived from the ICRG ratings by way of the World Bank's WIGs³¹—including the Control of Corruption indicator, good marks on which is a *sine qua non* for MCC eligibility.³²

Freedom in the World ratings, Freedom House

Freedom House is a non-profit NGO that was founded in 1941 to bolster popular support for United States involvement in World War II.³³ In 1972, it began to publish the Freedom in the World rankings, an annual survey of global political rights and civil liberties.³⁴

The Freedom in the World rankings assign countries numerical ratings between 1 and 7 for each of two dimensions, political rights and civil liberties, with 1 representing the most freedom and 7 the least.³⁵ In 2010, Freedom House produced political rights and civil liberties ratings for 194 countries and 14 territories.³⁶

²⁸ The MCC is a US aid initiative that was established in January 2004. MCC, *About MCC*, <<http://www.mcc.gov/mcc/about/index.shtml>> (accessed August 5, 2011) (hereinafter *About MCC*). Descriptions of the MCC's methodology in this paper focus on the approach utilized in Fiscal Year 2011 and, to some extent, FY 2012; in late 2011, the MCC introduced a somewhat different methodology for partial use in FY 2012 to which it will fully transition by FY 2013. MCC, *Report on the Criteria and Methodology for Determining the Eligibility of Candidate Countries for Millennium Challenge Account Assistance in Fiscal Year 2012* (2011), 4–6. The MCC identifies candidates for assistance based on the requirement that per-capita income levels fall beneath certain thresholds set by the World Bank. MCC, *Step 1: Identify Candidate Countries*, <<http://www.mcc.gov/pages/selection>> (accessed August 5, 2011). Under the FY 2011 methodology, it then selects countries eligible for assistance with reference to 17 indicators developed by third parties that are grouped into three broad categories: Ruling Justly, Investing in People, and Economic Freedom. MCC, *Selection Indicators*, <<http://www.mcc.gov/pages/selection/indicators>> (accessed August 5, 2011). A country must perform above the median in its income peer group—countries are sorted into Low Income and Lower Middle Income groups—on at least half the indicators in each of the three policy categories, and above the median on the Control of Corruption indicator found in the Ruling Justly category, in order to become eligible for assistance: MCC, *Guide to the MCC Indicators and the Selection Process* (2011), 2–3, available at <<http://www.mcc.gov/documents/reports/reference-2010001040503-fy11guidetotheindicators.pdf>> (accessed August 5, 2011) (hereinafter *Guide to the MCC Indicators*). As of August 2011, the MCC had approved \$7.4 billion in compacts with 23 countries. *About MCC* (above).

²⁹ *Guide to the MCC Indicators* (n. 28), 3–4.

³⁰ Kaufmann et al., "Governance Matters" (n. 27); 67. Of the ICRG political risk components, only Religious Tensions and Socioeconomic Conditions play no role in the WIGs. *Ibid.*

³¹ *Guide to the MCC Indicators* (n. 28), 11–20, 26–8.

³² *Ibid.*, 2–3.

³³ FreedomHouse.org, *Freedom House: A History*, <<http://freedomhouse.org/template.cfm?page=249>> (accessed August 5, 2011).

³⁴ FreedomHouse.org, *Freedom in the World: 2010 Edition, Introduction*, <http://www.freedomhouse.org/template.cfm?page=351&ana_page=362&year=2010> (accessed August 5, 2011).

³⁵ *Ibid.*

³⁶ *Ibid.*

Freedom House asserts that its ratings are used by “policymakers, the media, international corporations, civic activists, and human rights defenders to monitor trends in democracy and track improvements and setbacks in freedom.”³⁷ The ratings appear frequently in the media; a LexisNexis search of major newspapers and magazines during 2008 yields almost 300 articles referencing Freedom House.³⁸ Like the ICRGs, the Freedom in the World ratings are also popular with academics; many scholarly papers treat the ratings as proxies for freedom, governance, or democracy.³⁹

The Freedom in the World ratings are also used by the MCC as criteria for the disbursement of funds to developing countries, with two of the 17 MCC threshold indicators furnished directly by Freedom House’s political rights and civil liberties ratings.⁴⁰ As noted above, the ICRGs, as incorporated into the World Bank WIGs, play a prominent role in determining MCC eligibility, but Freedom House’s indicators arguably play an even more critical role in MCC selectivity. First, Freedom House’s ratings act as stand-alone threshold indicators; unlike the measures included in the WIGs, they are not aggregated with other measures before playing a role in selecting MCC recipients.⁴¹ Second, the quantized nature of the Freedom House ratings increases their importance. As Steven Radelet explains, “there are only seven possible scores” on the Freedom House indicators, and “[s]ince many countries are assigned exactly the same score (e.g., a 4 or 5), they are bunched together around the median score, which is where the [MCC] draws the line between passing or failing.”⁴² As a result, small variations in Freedom House’s indicator levels can have large effects on candidate country eligibility.

³⁷ Freedom in the World (n. 34).

³⁸ A LexisNexis search of Major Newspapers and Magazine Stories from January 1, 2008 to December 31, 2008 yields 299 results for “Freedom House,” but nets only 10 results for “International Country Risk Guide.”

³⁹ Academics have used the Freedom House ratings to investigate such phenomena as the effect of democracy on health, Álvaro Franco et al., “Effect of Democracy on Health: Ecological Study,” *British Medical Journal* 329 (2004): 1421; the relationship between governance and the efficacy of government projects, Jonathan Isham et al., “Civil Liberties, Democracy, and the Performance of Government Projects,” *World Bank Economic Review* 11 (1997): 219; and the causal interplay between economic freedom, political freedom, and economic growth, W. Ken Farr et al., “Economic Freedom, Political Freedom, and Economic Well-Being: A Causality Analysis,” *Cato Journal* 18 (1998): 247. More often than the ICRG ratings, moreover, the Freedom House indicators are used in academic articles not just as elements in a regression, but to summarize worldwide or regional changes in freedom, much as the ratings are used in journalism. See, e.g., Fareed Zakaria, “The Rise of Illiberal Democracy,” *Foreign Affairs* (November–December 1997): 22, 23–4 (observing that according to Freedom House’s Freedom in the World survey, “half of the ‘democratizing’ countries in the world today are illiberal democracies”); Larry Diamond, “Thinking About Hybrid Regimes,” *Journal of Democracy* (April 2002): 21, 26 (corroborating a “steady overall rise in freedom in the world” by noting that “the average score on the combined seven-point Freedom House scale improv[ed] from 4.47 in 1974 to 3.47 in 2001”).

⁴⁰ Guide to the MCC Indicators (n. 28), 3.

⁴¹ Ibid.

⁴² Radelet, “Challenging Foreign Aid” (n. 26), 36.

Minorities at Risk dataset, MAR Project

The Minorities at Risk (MAR) Project was founded in 1986 by Ted Robert Gurr, and moved to the University of Maryland along with Gurr in 1988.⁴³ According to its website, the Project “tracks 283 politically active ethnic groups throughout the world from 1945 to the present—identifying where they are, what they do, and what happens to them.”⁴⁴

MAR data includes both qualitative risk assessments and quantitative data comprising approximately 400 variables, including 71 variables identified as “core”⁴⁵ that code characteristics ranging from “Government repression of group”⁴⁶ to “Group organization and representation”⁴⁷ to “Group concentration.”⁴⁸ While the qualitative risk assessments describe risk without assigning it a numerical value,⁴⁹ the quantitative data take complex phenomena and reduce them to numbers, thus functioning as indicators.⁵⁰

The MAR Project is “designed to provide information in a standardized format that will aid comparative research and contribute to the understanding and peaceful accommodation of conflicts involving communal groups,” and to that end, the Project makes its materials available to “researchers, students, public officials, journalists, activists, and other interested officials.”⁵¹ However, the most common use of the dataset of indicators appears to be academic. An MAR webpage listing publications making use of the dataset between 1993 and 2004 identifies 71 publications,⁵² while a 2004 paper by Joanne Manrique finds 85 articles referring to the dataset.⁵³

In contrast, although journalists are an intended audience of the MAR Project, journalistic uses for the indicator set seem to be quite limited. A LexisNexis search of major newspapers and magazines from 2008 yields not one non-academic article

⁴³ MAR, About MAR, <<http://www.cidcm.umd.edu/mar/about.asp>> (accessed August 5, 2011).

⁴⁴ Ibid.

⁴⁵ Minorities at Risk, *Minorities at Risk (MAR) Codebook 2–3* (February 2009), available at <http://www.cidcm.umd.edu/mar/data/mar_codebook_Feb09.pdf> (accessed August 5, 2011) (hereinafter MAR Codebook).

⁴⁶ Ibid., 23.

⁴⁷ Ibid., 12.

⁴⁸ Ibid., 7.

⁴⁹ See, e.g., MAR, Data, Assessment for Hazaras in Afghanistan, <<http://www.cidcm.umd.edu/mar/assessment.asp?groupId=70001>> (accessed August 5, 2011) (noting that “[t]he likelihood of Hazara protest is moderate” but “[t]hey are unlikely to rebel in the future”); MAR, Data, Assessment for Mohajirs in Pakistan, <<http://www.cidcm.umd.edu/mar/assessment.asp?groupId=77007>> (accessed August 5, 2011) (explaining that “Mohajirs [*sic*] risk for rebellion is moderate,” and that “Mohajirs have a history of persistent protest”).

⁵⁰ See MAR Codebook (n. 45), 1, 6–24 (describing the variables as “quantitative indicators”).

⁵¹ Christian Davenport, “Minorities at Risk,” *Dataset Users’ Manual* vol 5 (2003) available at <http://www.cidcm.umd.edu/mar/margene/mar-codebook_040903.pdf> (accessed August 5, 2011).

⁵² MAR, Resources, Publications, <<http://www.cidcm.umd.edu/mar/publications.asp>> (accessed August 5, 2011).

⁵³ Joanne Manrique, “Uses of the Minorities at Risk (MAR) Variables” (March 2004) 3.

referring to the Project.⁵⁴ The paucity of journalistic mentions of the Project suggest that advocacy and public policy uses of the dataset are similarly restricted, since NGOs appear not to be using its data to publicize practices in their own countries, and states seem not to employ its data to place public pressure on other states that treat minorities badly. Indeed, outside of academic articles and MAR's own website and associated sites, the dataset has little online presence.⁵⁵ It is certainly possible that activists and public officials are using the MAR indicators to identify problem areas and focus resources and attention on those areas without referring to the indicators publicly. Nonetheless, the most prevalent use of the MAR indicators seems to be academic.

Evaluating the Demand Hypothesis

The test of the Demand Hypothesis first predicts the levels of accountability each indicator would display were the Demand Hypothesis to hold true. It then compares these predictions to the levels of accountability observed for each indicator.⁵⁶

The predictions of the Demand Hypothesis

If the Demand Hypothesis were valid, one would expect the ICRGs to display the highest level of accountability, the Freedom in the World rankings somewhat lower levels, and the Minority at Risk variables the lowest levels. After all, the ICRGs are used to determine investment flows and MCC development funding allocations; private investment flows are large in magnitude, even for developing countries,⁵⁷

⁵⁴ A LexisNexis search of Major Newspapers and Magazine Stories from January 1, 2008 to December 31, 2008 produces only four results for "Minorities at Risk" AND ("Gurr" OR "Maryland" OR "CIDCM").

⁵⁵ But see, e.g., <<http://www.ecoi.net>>, Iran, Human Rights Issues, Turkmen, <<http://www.ecoi.net/189469::iran/328787.321470.8956...mr.321539/turkmen.htm>> (accessed August 5, 2011) (quoting MAR assessments with the aim of providing "country of origin information" to "asylum lawyers, refugee counsels and persons deciding on claims for asylum and other forms of international protection").

⁵⁶ Our analysis will largely focus on accountability "on the books"—on the amounts of transparency, participation, review, and reason-giving suggested by publicly available documentation regarding each indicator. This approach does not take into account informal accountability—non-proceduralized, spontaneous responses to information requests, attempts to participate, and complaints about indicator levels—in part because it is difficult to observe. Moreover, such accountability may be less effective than formal, proceduralized accountability, since it depends on outsiders taking the initiative to seek and generate accountability where there appears to be none; as we note later, selectively provided accountability may also undermine accountability in a general sense, particularly when such selective accountability is selectively furnished. See below nn. 106, 122.

⁵⁷ A 2005 report noted that "in the 1970s bank lending was nearly 20 times the bond issues for emerging markets, but by the 1990s bond issues had surpassed bank lending." Facundo Martin et al., The World Bank Group, "The Demand for Loans," *Public Policy for the Private Sector* (April 2005), 1, available at <<http://rru.worldbank.org/documents/publicpolicyjournal/289martin.pdf>> (accessed August 5, 2011).

and MCC grants can amount to hundreds of millions of dollars. The Freedom House rankings play little role in directing investment flows, but they are a critical component of the MCC criteria for allocating funding. Moreover, their prominence in journalism is notable, since coverage of regimes by the media can affect the reputations of countries, political influence, prestige within the community of world leaders, and even investment flows. Consequently, the stakes involved in the use of the Freedom House ratings are significant, if not so high as those involved in decisions made based on the ICRGs. The MAR variables, on the other hand, used as they are almost exclusively in scholarship, cannot be considered to affect high-stakes decisions.

The methodologies of the test cases

To test these predictions, it is necessary to first investigate the methodologies employed in generating each of our test case indicators. It is then possible to judge the extent to which these methodologies offer the elements of accountability: participation, transparency, reason-giving, and review.⁵⁸

The ICRG political risk ratings are subdivided into components and subcomponents, with an analyst assigning each country a value for each of these constituents.⁵⁹ These assessments are based wholly on the subjective judgment of analysts; as ICRG documentation explains, “the ICRG staff collects political information” and “political risk assessments are made on the basis of subjective analysis” of this information, though “[t]o ensure consistency, both between countries and over time, points are assigned by ICRG editors on the basis of a series of pre-set questions.”⁶⁰ The points assigned are then added to yield an aggregate score and a risk assessment.

As for Freedom House, its ratings are developed using the services of 50 in-house and consultant analysts and 18 “senior-level” academic advisers.⁶¹ To begin, an analyst writes a report describing the situation within each country or territory⁶² and then calculates preliminary ratings by answering a series of diagnostic questions under the umbrella of either “political rights” or “civil liberties” and assigning point values to each response.⁶³ As Freedom House’s website explains, “analysts use[] a

⁵⁸ We will consider legality as a component of accountability only where there is some reason to believe a generator is not following its own publicly declared processes.

⁵⁹ ICRG Methodology Manual (n. 9), 2–7.

⁶⁰ *Ibid.*, 2.

⁶¹ Freedom House, *Freedom in the World: 2010 Edition*, Methodology, <http://www.freedomhouse.org/template.cfm?page=351&cana_page=363&year=2010> (accessed August 5, 2011) (hereinafter *Freedom House Methodology*).

⁶² *Ibid.*

⁶³ *Ibid.* The political rights rating is based on a checklist of 10 questions, each worth a maximum of four points, and two additional “discretionary” questions that can add or subtract points from a country’s aggregate raw score; the civil liberties rating is based on 15 questions that are each weighted four points. *Ibid.* Thus, a country may be awarded as many as four points toward its civil liberties raw score based on the answer to the question, “Is there open and free private discussion”? Freedom House has developed a set of sub-questions to guide the inquiry set up by each diagnostic question, though an

broad range of sources of information—including foreign and domestic news reports, academic analyses, nongovernmental organizations, think tanks, individual professional contacts, and visits to the region—in preparing the . . . reports” and answering the questions.⁶⁴

After each question has been answered and scored, Freedom House’s analysts aggregate these values to produce “raw scores” that correspond to ratings. Depending on the raw scores, countries may be rated “Free,” “Partly Free,” or “Not Free.”⁶⁵ Analysts answer questions and assign scores using as a benchmark the points assigned in the previous edition of *Freedom in the World*; changes in raw points are made only “if there has been a real world development during the year that warrants a change.”⁶⁶ Moreover, once analysts have assigned raw points to countries, Freedom House puts the preliminary ratings that result through a multi-stage internal review process.⁶⁷ Any major proposed numerical shifts are subject to intensified investigation in these reviews.⁶⁸

The MAR Project develops indicators only for “Minorities at Risk,” or ethno-political groups that (1) “collectively suffer[], or benefit[] from, systematic discriminatory treatment vis-à-vis other groups in a society”; and (2) “collectively mobilize [] in defense or promotion of [their] self-defined interests.”⁶⁹ The Project determines which such groups to track according to a set of relatively complex criteria.⁷⁰ As already noted, the MAR quantitative dataset includes about 400 variables.⁷¹ These variables are coded primarily by “graduate and undergraduate students who . . . undergo[] a rigorous training period,” with “all coding . . . reviewed by senior editors and by the research director.”⁷² Student coders are guided in their

analyst is not expected to consider every sub-question in assigning a rating to each country. The question quoted above is to be answered by reference to two sub-questions: (1) “Are people able to engage in private discussions, particularly of a political nature (in places including restaurants, public transportation, and their homes) without fear of harassment or arrest by the authorities?”; and (2) “Does the government employ people or groups to engage in public surveillance and to report alleged antigovernment conversations to the authorities?”: Freedom House, *Freedom in the World: 2010 Edition, Checklist Questions and Guidelines*, <http://freedomhouse.org/template.cfm?page=351&ana_page=364&year=2010> (accessed August 5, 2011).

⁶⁴ Freedom House Methodology (n. 61), 60.

⁶⁵ *Ibid.*

⁶⁶ *Ibid.*

⁶⁷ Ratings are first reviewed individually; then they are re-examined on a comparative basis in “regional meetings” involving analysts, academic advisors with regional expertise, and Freedom House staff. Finally, they are scrutinized on a cross-regional basis to ensure comparability and consistency. *Ibid.*

⁶⁸ *Ibid.*

⁶⁹ MAR Codebook (n. 45), 1.

⁷⁰ These criteria include whether (1) “[m]embership in the group is determined primarily by descent by both members and non-members”; (2) “[m]embership in the group is recognized and viewed as important by members and/or non-members”; (3) “[m]embers share some distinguishing cultural features”; and (4) “[t]he group has at least 100,000 members or constitutes one percent of a country’s population.” *Ibid.*, 1–2. It is unclear, from the MAR documentation, who determines whether a group qualifies for inclusion in the dataset, but the documentation does explicitly observe that “the Project does not make claims regarding the comprehensiveness of the dataset.” *Ibid.*, 2.

⁷¹ *Ibid.*, 5–24.

⁷² *Ibid.*, 3.

efforts by the labels affixed to the codes that are assigned to each variable,⁷³ but it is unclear whether coders are also given some more detailed definition of the variables they code. Some of the information sources relied upon are listed along with the qualitative assessments,⁷⁴ and the MAR documentation explains that “[s]elected hard-copy source materials” are kept in the Project archives and are available to individual researchers “by arrangement with the Project coordinator.”⁷⁵ All coding “is conducted using open-source information.”⁷⁶

Observed levels of accountability

On the basis of the information provided by the PRS Group, Freedom House, and the MAR Project, we can identify the levels of participation, transparency, reasoning, and review that each indicator generator furnishes.

Participation

ICRG documentation does not suggest that any systematic provision has been made in its process for consultation with outsiders.⁷⁷ Groups wishing to provide input may avail themselves of the sales contact information provided on the PRS Group website, but no contact information accompanies the website’s listing of the Group’s editorial board.⁷⁸

The Freedom in the World methodology does not suggest that any formal provision is made in the ratings process for participation by outside actors. However, the “broad range of sources of information” used by analysts to assign ratings includes “foreign and domestic news reports, academic analyses, nongovernmental organizations, think tanks, individual professional contacts, and visits to the region,” the latter four of which have the potential to produce opportunities for participation by affected actors.⁷⁹ Freedom House provides lists of staff, analysts, and academic advisers, but contact information for these persons is unavailable:⁸⁰

⁷³ For example, “Restrictions on religion” may take the values 0, 1, 2, 3, and –99, which correspond to “No restrictions,” “Activities informally restricted,” “Activity somewhat restricted,” “Activity sharply restricted,” and “No basis for judgment,” respectively. *Ibid.*, 11.

⁷⁴ See, e.g., MAR, Data, Assessment for Pashtuns in Afghanistan, <<http://www.cidcm.umd.edu/mar/assessment.asp?groupId=70002>> (accessed August 5, 2011) (citing reports and chapters from the International Crisis Group, *New York Times*, and *Washington Post*, among other sources, as references).

⁷⁵ MAR Codebook (n. 45), 4.

⁷⁶ *Ibid.*, 3.

⁷⁷ ICRG Methodology Manual (n. 9), 2 (describing the process whereby ICRG ratings generally, and political risk ratings in particular, are determined, and omitting to describe any participatory step).

⁷⁸ The PRS Group, Contact Us, <<http://www.prsgroup.com/ContactUs.aspx>> (accessed August 5, 2011).

⁷⁹ Freedom House Methodology (n. 61).

⁸⁰ Freedom House, Freedom in the World: 2010 Edition, Survey Team, <http://www.freedomhouse.org/template.cfm?page=351&ana_page=366&year=2010> (accessed August 5, 2011).

the Freedom House website provides only general phone numbers and email addresses for each of its offices.⁸¹

Unlike the other two indicators studied here, the MAR Project makes some provision for participation in the generation of its quantitative variables. MAR Project documentation encourages researchers “to carry out their own consistency and validity checks on indicators they use or adapt from the MAR dataset” and notes that “Project staff would greatly appreciate being appraised of the results of such analysis.”⁸² The MAR website identifies Project staff and members of its advisory board, and provides email addresses for these persons to facilitate communication.⁸³ Finally, the Project has reached out to its users, convening “a workshop at the [American Political Science Association] conference in Philadelphia in September 2006 to bring together a group of scholars to help MAR effectively deal with selection bias.”⁸⁴ As a result, MAR has developed a strategy that it “aim[s] to implement over the coming years” that will involve the “develop[ment of] new, far more inclusive criteria to identify communal groups around the world for inclusion in the MAR dataset” and a “reduc[tion of] the number of variables coded by MAR in future updates.”⁸⁵ The MAR Project thus both enables and proactively seeks consultation with academic users of its indicators.

Transparency

The PRS Group publishes relatively detailed documentation describing how its ICRG ratings are calculated, specifying the components of the political risk rating, their weights, and the questions used to assign scores.⁸⁶ The Group also provides the names of the members of its editorial board, although these names are unaccompanied by other descriptive information.⁸⁷ The Group fails, however, to make public the sources of information that it draws on to make its assessments⁸⁸ and does not explain how these sources of information are chosen, or whether safeguards have been instituted to ensure their quality. The Group also does not

⁸¹ Freedom House, Contact Us, <<http://www.freedomhouse.org/content/contact-us>> (accessed April 12, 2012); FreedomHouse.org, Staff Directory, <<http://www.freedomhouse.org/content/our-staff>> (accessed April 12, 2012).

⁸² MAR Codebook (n. 45), 4.

⁸³ MAR, About MAR, Project Staff, <<http://www.cidcm.umd.edu/mar/about.asp#staff>> (accessed August 5, 2011).

⁸⁴ MAR, MAR and Selection Bias, <<http://www.cidcm.umd.edu/mar/announcement.asp?announcementid=26>> (accessed August 5, 2011).

⁸⁵ Ibid.

⁸⁶ ICRG Methodology Manual (n. 9).

⁸⁷ The PRS Group, Contact Us, <<http://www.prsgroup.com/ContactUs.aspx>> (accessed August 5, 2011).

⁸⁸ The PRS Group does state generally that, regarding the ICRGs, it “produces the information and data on which the ratings for the individual risk components are determined, together with its interpretation of that information or data.” ICRG Methodology Manual (n. 9), 2. But it is unclear whether this information includes a detailed list of the references employed in generating the political risk ratings, and in any case, this information is not available to non-purchasers of the ICRGs.

make public the names and qualifications of the staff analysts who do most of the work in calculating the ICRGs.

Freedom House publishes extensive information about all phases of its ratings-generation process for Freedom in the World, from enumerating the questions used to guide the assignment of raw scores⁸⁹ to describing the stages of review through which initial ratings must pass.⁹⁰ It also provides names and detailed biographies for both the analysts and academic advisers who work on each edition of the rankings, identifying the region to which each was assigned.⁹¹ Finally, Freedom House makes public a list of nearly 300 “Selected Sources” on which it draws to produce its ratings.⁹² Short of opening its deliberations to public scrutiny, then, Freedom House does nearly everything that could be expected to make the process by which it generates ratings transparent.

The MAR Project documentation provides few details about its coding process. The Project does not list the students who have the actual responsibility for coding its variables, so that the number, qualifications, and specializations of these students remain unclear. Moreover, while some review of student coding apparently occurs, the MAR documentation leaves ambiguous whether this review is mandatory or selective, multi-stage or single-stage, committee-based or carried out by individual staff.⁹³ As regards methodology, then, the Project fails to provide much transparency.

Reason-giving

The PRS Group accompanies its ratings with descriptive “analyses of events that affect the risk ratings,” which could be considered to be an exercise in reason-giving—but the Group supplies such analyses only for 20–25 important countries, out of the 140 total countries rated.⁹⁴ The Group does “provide[] not only the risk ratings for the countries it covers, but also the political information and financial and economic data on which those ratings are based” in order to make it “possible for the user to check through the information and data so as to assess the ratings given against his or her own assessments or against some other risk rating system.”⁹⁵ Such information could be considered to be either a terse example of reason-giving or an additional source of transparency.

Freedom House provides detailed “country reports” for every state rated in Freedom in the World.⁹⁶ Each report contains a “historical background and a

⁸⁹ Freedom in the World: 2010 Edition (n. 61), Checklist Questions and Guidelines.

⁹⁰ Freedom House Methodology (n. 34).

⁹¹ Freedom House, Freedom in the World: 2010 Edition (n. 34), Survey Team.

⁹² FreedomHouse.org, Freedom in the World: 2010 Edition, Selected Sources, <<http://www.freedomhouse.org/report/freedom-world-2010/selected-sources>> (accessed April 12, 2012).

⁹³ MAR Codebook (n. 45), 3–4.

⁹⁴ The PRS Group, International Country Risk Guide (ICRG), <<http://www.prsgroup.com/ICRG.aspx>> (accessed August 5, 2011).

⁹⁵ ICRG Methodology Manual (n. 9), 16.

⁹⁶ FreedomHouse.org, 2010 Edition: Country Reports, <<http://www.freedomhouse.org/template.cfm?page=21&year=2010>> (accessed August 5, 2011).

brief description of the year's major developments, as well as a section summarizing the current state of political rights and civil liberties."⁹⁷ These reports can be viewed either as information meant to complement the ratings or as reason-giving exercises intended to justify the ratings.

Finally, the MAR Project offers "qualitative assessments" for each ethnopolitical group analyzed that contain both a "risk assessment" and an "analytic summary"; the risk assessment "summarizes whether the group is at risk of rebellion, protest, or repression," while the summary "gives a brief history of the group and its relations with the state."⁹⁸ The risk assessment is usually too terse to justify considering it to be an exercise in reason-giving, with assessments occupying fewer than five sentences in most cases.⁹⁹ The analytic summary, on the other hand, is an approximately 1,000-word description of political and historical conditions that—by noting the values assigned to major related variables in parentheses following relevant sentences—does serve as a brief justification for 10–15 of the most salient indicators.¹⁰⁰

Review

None of the three indicators is subject to external review procedures. There appears to be no provision for third-party review of disputed ICRG ratings. As for Freedom House, while ratings pass through several stages of review by staff and consultants working for Freedom House, there does not appear to be any institutionalized pathway for outsiders to appeal disputed ratings.¹⁰¹ Finally, while the MAR Project exercises oversight over student coding,¹⁰² there is certainly no review of such coding by independent, impartial arbiters at the request of interested parties.

Comparing predictions to observations

The ICRGs, then, offer moderate levels of transparency in their generation, low levels of participation, moderate levels of reason-giving, and low levels of review. Freedom House provides high levels of transparency, moderate levels of participation, high levels of reason-giving, and low levels of review. The MAR Project furnishes relatively low levels of transparency, high levels of participation, moderate levels of reason-giving, and low levels of review. It is hard to synthesize these component levels into unified judgments about how accountable each indicator is in its generation. But it seems clear that the ICRGs, which appear to have the most consequential uses, do not display greater accountability than the Freedom

⁹⁷ Freedom House Methodology (n. 61).

⁹⁸ MAR, Data, <<http://www.cidcm.umd.edu/mar/data.asp>> (accessed August 5, 2011).

⁹⁹ See, e.g., MAR, Data, Assessment for Roma in Serbia, <<http://www.cidcm.umd.edu/mar/assessment.asp?groupId=34506>> (accessed August 5, 2011) (offering a six-sentence Risk Assessment).

¹⁰⁰ See, e.g., *ibid.* (including a 931-word Analytic Summary).

¹⁰¹ Freedom House Methodology (n. 61).

¹⁰² MAR Codebook (n. 45), 3.

House ratings or the MAR variables. The Demand Hypothesis, in short, fails to explain observed accountability.

Explaining the failures of the Demand Hypothesis

Why does the Demand Hypothesis do such a poor job of predicting the accountability levels provided by indicator generators? Closer investigation of the Hypothesis reveals a few potential flaws. To begin, it seems to fail to account for how indicator users' behavior is motivated not by the stakes of indicator-based decisions, but by their interest in these decisions. It also appears to neglect how indicator targets' actions are limited by lack of resources, inability to exit, and collective action problems. Finally, the Hypothesis likely overestimates generators' willingness to furnish increased accountability to outsiders.

Users: Not stakes, but interest

The Demand Hypothesis uses the stakes of decisions made based on indicators to determine how likely users should be to demand improved indicator accuracy. It thus assumes that when decision-making stakes are high, interest in indicator accuracy should be high as well. While this stakes-interest correlation may obtain in most circumstances, it can break down in the context of governance indicators for at least three reasons.

Hedging by users

Even if investors and development agencies have a strong interest in ensuring that high-stakes decisions are based on accurate information, both types of users can hedge against the risk that any given source of information will be inaccurate by using multiple sources of information. The resort to supplementary information reduces the demand for accuracy of any given indicator. This appears to be particularly true with respect to the ICRGs. As Mikelle A. Calhoun notes, "risk ratings do not seem to be a significant source of information for firms and reliance on them is limited at best"; instead, "[a]vailable information reflects that firms evaluate country risk using methods that range from highly complex econometric models to purely qualitative, judgmental approaches."¹⁰³ This suggests that investors may not base their investment decisions even in large part on commercial risk ratings like the ICRGs, but rather may use these ratings merely to supplement their own in-house analyses. As for the MCC, many of its criteria rely not solely on measures produced by single institutions like the PRS Group, but on aggregate

¹⁰³ Mikelle A. Calhoun, "Seeking Meaningful Country-Risk Information: A Wealth of Information, a Void in Understanding," *Perspectives on Strategic Risk Management* (Torben Juul Andersen ed., CBS Press, 2006), 147, 154.

indicators compiled by the World Bank that combine measures from as many as 20 other sources.¹⁰⁴ In the case of the ICRGs, then, we should not be surprised to see only moderate levels of accountability, which suggests few demands from users for increased accuracy or accountability; for really high-stakes decisions, users may have diversified their risk and consequently have less interest in ICRG accuracy.

Insulation from consequences of decisions

The hedging argument seems to apply less to Freedom House's ratings than the ICRGs, though. If five of the 17 indicators used by the MCC are furnished by the World Bank Institute in the form of aggregate measures, the other 12 come from single institutions, including the Civil Liberties and Political Rights criteria taken directly from Freedom House's Freedom in the World ratings.

In this context, though, the wide variety of factors motivating organizations like the MCC can decrease their interest in indicator accuracy, even when decision-making stakes are high. Oversight mechanisms exist to ensure that the MCC channels its funds toward countries that can effectively use them, most notably in the form of congressional review of MCC activities. Since MCC budgets are set by Congress,¹⁰⁵ if these legislators have an interest in ensuring that funds go to well-governed countries, and if they can ascertain how efficiently funds are being allocated, then those working for the MCC should develop a strong interest in the accuracy of indicators used to direct allocations.

But neither of these conditions is necessarily met. Legislators may want aid to be spent effectively, but they also likely want aid to be used in support of other aims, including furthering US geopolitical goals and supporting domestic economic interests. Moreover, even ineffective aid disbursements may have utility in demonstrating, to constituents and other nations, a commitment to development. Finally, legislators will often have difficulty determining how well aid is being spent; a "lack of results" can plausibly be attributed to so many factors other than improper allocation of aid that legislators will usually be unable to monitor how well such funds are distributed. For all these reasons, the MCC, and other development agencies that might rely on indicators to make funding decisions, should not necessarily have a strong interest in the accuracy of these indicators, but rather merely an interest in justifying expenditures to those who exercise oversight. The two are, of course, not equivalent.

Unwillingness by users to undermine indicators

Even if investors, development agencies, and academics have a strong interest in the accuracy of the indicators on which they base their decisions, they may nonetheless avoid making public demands for improved accuracy or accountability because

¹⁰⁴ Guide to the MCC Indicators (n. 28), 3–4.

¹⁰⁵ See, e.g., Steven Radelet, "Will the Millennium Challenge Corporation Be Different?," *Washington Quarterly* 26 (2003): 183 (describing Congress's role in allocating MCA funds).

such demands risk discrediting measures to which they have linked their own reputations. By publicizing the inadequacies of political risk services like the PRS Group, investors risk alerting backers to the weakness of their own strategies. Development agencies similarly risk sanction from legislative oversight committees and protest from country candidates for aid, and academics open themselves to the charge that their work is invalid. The users of governance indicators should thus generally seek either to lodge any criticisms of accuracy or accountability levels in private, or not to make these criticisms at all.¹⁰⁶

Targets: Not interest, but capacity

One defect of the Demand Hypothesis, then, appears to be that even if users' interest in indicator accuracy determines the intensity of their demands for accuracy and accountability, the stakes of users' indicator-based decisions are a poor proxy for such interest. We can identify another possible defect of the Hypothesis, regarding the behavior of indicator targets. Even if targets have an interest in ensuring that indicators are accurate—or in seeking opportunities to influence how indicators are set—when important decisions are based on those indicators, they may still lack the capacity to influence indicator generators' behavior for at least three reasons.

Lack of credibility or means

When indicator targets decide whether to demand improved accuracy or accountability from indicator generators, they almost certainly take into account the likelihood that their demands will be heeded. After all, making demands may be costly in terms of time, human capital, and deflection of lobbying resources from other activities. We can identify a few factors that might discourage targets from making such demands even if they have a strong interest in changing indicator accuracy or accountability.

¹⁰⁶ We should not be surprised if these criticisms are being made in private, and if indicator generators are actually responding by offering special levels of transparency, participation, reason-giving, and review to select users, again in private. Thus, a survey of Freedom House's Board of Trustees suggests that nearly half of the organization's leadership have spent a substantial portion of their careers working for the federal government. See Freedom House, Board of Trustees, <<http://www.freedomhouse.org/content/our-leadership>> (accessed April 12, 2012) (providing a list of the organization's trustees that links to biographies of each). According to their biographies, of the 39 trustees listed, 18 have had significant experience in government. The proportion of Freedom House staff with significant government experience is likely even higher. This is as should be expected for a prominent NGO analyzing freedom levels abroad, but it also indicates that US government agencies may have means of holding Freedom House to account in its indicator generation that are not apparent from publicly available descriptions of indicator methodology. While such selectively and, to some extent, secretly elevated levels of transparency, participation, reason-giving, and review bolster accountability to some extent, they also undermine accountability by reducing the legality of indicator generators. Such selective access leaves other actors less certain that the indicator generator is actually relying on the methodology that its public statements would indicate.

Those who use indicators to judge the quality of governance in a country likely do so in part because the indicator generators are ostensibly impartial observers, in contrast to data supplied either by the users' own governments or the businesses and governments of target countries. Indicator generators should thus be reluctant to heed the demands of targets; such interactions might damage their reputations for objectivity, especially if the targets represent countries that have reputations for oppression and secrecy.

Target countries may also lack the means to contact and persuade generators to change the way they develop indicators. Countries with insufficiently developed information management systems will be at a disadvantage in gathering and presenting data that challenges the results encapsulated in an indicator. Moreover, if a country has few persons that have been educated abroad, or that are employed in prominent universities, international companies, or development agencies, it will be hard for that country to open channels of communication with indicator generators. Finally, any lobbying effort requires resources—to fund research, travel, the retention of intermediaries, and so on. Less wealthy countries will thus be less able to lodge demands for accountability.

Inability to marshal credible threats of exit

The likelihood that demands for enhanced accuracy or accountability will be successful is affected not only by the credibility and means of indicator targets, but by their capacity to marshal credible threats of exit from the indicator-target relationship. As Albert O. Hirschman has explained, when an organization fails to meet performance expectations, customers or members have two routes through which they can alert management to its failings: exit and voice.¹⁰⁷ If exit is easy, then customers or members are unlikely to expend the effort that voice requires; if exit is impossible, however, organizational management becomes unlikely to heed any complaints that are voiced.¹⁰⁸ We can therefore “spell out the conditions under which voice (a) will be resorted to and (b) bids fair to be effective: there should be the possibility of exit, but exit should not be too easy or too attractive as soon as deterioration of one's own organization sets in.”¹⁰⁹

Indicator targets cannot simply choose to stop being measured and therefore exit their relationship with indicator generators. They could, however, refuse to cooperate with the generators by denying researchers visas and declining to respond to any requests for information. Indicator targets can also attempt to induce the users of indicators to stop using indicators to evaluate the target, which would constitute a type of “effective exit” even as measurement and indicating activities by the generator continued.

¹⁰⁷ See generally Albert O. Hirschman, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Harvard University Press, 1970).

¹⁰⁸ *Ibid.*, 55.

¹⁰⁹ *Ibid.*, 83.

These exit methods are not likely to be effective, however, and hence lack credibility. By refusing to cooperate with indicator generators' efforts to gather information, targets surrender any control they might have over the character of this information. Moreover, as we have already discussed, targets are unlikely to be viewed by users as credible evaluators of their own governance. Efforts by targets to implement an "effective exit" by inducing users to switch from one indicator to another might foster perceptions that the favored indicators are actually in thrall to the targets, while the disfavored indicators are impartial and accurate.

The "effective exit" approach also includes many of the disadvantages that usually characterize "voice" options and the absence of which makes "exit" options attractive. Convincing users to stop using a given indicator, or to switch to another indicator, represents a public good for all targets dissatisfied with the given indicator, and hence gives rise to the collective action problems that we will consider in the following section. Moreover, the process of persuading users to change their behavior is likely to be a costly one, particularly if such users have publicly committed to rely on a set of indicators.¹¹⁰

The inability of targets to exit from their relationship with indicator generators, even in an "effective" sense, should make their demands for accuracy or accountability far less powerful. Consequently, targets should be less likely to voice such demands.

Barriers to collective action

As Mancur Olson has explained, an individual member of a large organization can rarely "have a noticeable effect on the situation of his organization, and he can enjoy any improvements brought about by others whether or not he has worked."¹¹¹ On the other hand, "in some small groups each of the members, or at least one of them, will find that his personal gain from having the collective good exceeds the total cost of providing some amount of that collective good."¹¹² Thus, "the larger the group, the farther it will fall short of providing an optimal amount of a collective good,"¹¹³ though "[i]n small groups with common interests there is . . . a surprising tendency for the 'exploitation' of the great by the small."¹¹⁴ Applying this model to indicator generation, if some targets find their welfare significantly affected by decisions based on indicators, and these targets can, through their own efforts, secure enhanced accuracy or accountability from indicator generators, such "large" targets

¹¹⁰ Thus, the MCC cannot simply choose to abandon the WGI or the Freedom in the World ratings because some indicator targets prevail upon them to do so; many potential candidates for aid have relied upon the MCC's commitments to use these indicators in creating their own development strategies.

¹¹¹ Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard University Press, 1971), 16.

¹¹² *Ibid.*, 33–4.

¹¹³ *Ibid.*, 34–5 (emphasis omitted).

¹¹⁴ *Ibid.*, 35 (emphasis omitted).

should demand accuracy and accountability. But do such auspicious circumstances for overcoming barriers to collective action exist?

As we have already explained, the ICRGs are used in determining investment flows and aid allocations, and Freedom House's Freedom in the World rankings are employed in aid distribution and journalism. We would expect targets to have a significant interest in how such consequential indicators are set.

We would also expect some indicator targets to be "larger" than others in terms of their capacity to influence indicator generators. Some target countries are much wealthier or represent much larger economies than others. They should thus have more resources and influence to draw on in lobbying efforts, so that if the ability to bring about changes in accuracy or accountability levels in indicators depends on lobbying capacity, several targets will be substantially "larger" than the rest. Whether any of these larger targets could have the capacity, by itself, to bring about a change in indicator-generating methodologies is not quite clear, but such a prospect does not seem impossible.

Whether or not a single large target could prevail upon an indicator generator to change its processes, however, ultimately matters little for our collective action analysis, because a serious dilemma faces target countries seeking to influence indicator generation: those countries with the *greatest* capacity to prevail upon indicator generators to change the way that they set indicators will also be the countries with the *least* interest in their accuracy or accountability. After all, the wealthiest countries with the largest economies will (1) probably be independently analyzed by investors without much reference to relatively simple indicators like the Risk Guides; (2) depend little on the comparatively small sums of money offered by organizations like the MCC; and (3) be the subjects of a great deal of journalistic coverage, entirely apart from that which relies on Freedom House ratings and their ilk. The countries with the most to lose from unfavorable indicators, in contrast, will be least able to convince generators to change their ways.

Even though some indicator targets should have a significant interest in influencing indicator generation, and some targets will likely be quite "large," we should ultimately expect neither "large" nor "small" targets to make demands upon generators for improved accuracy or accountability. Thus, if demands are not quieted by targets' lack of means and credibility, or by the absence of the credible threats of exit needed to give force to their demands, collective action problems should nonetheless accomplish the same result.

Generators: Reluctance to bolster accountability

Finally, even if we ignore the problems of the Demand Hypothesis in predicting user or target behavior, it nonetheless remains true that indicator generators may not respond to demands from such actors for improved accuracy by boosting accountability. While improved accountability may seem to have a unique capacity to bolster accuracy, increasing accountability may also tend to intensify demands for influence by outsiders, which indicator generators should want to avoid. As a

consequence, indicator generators may choose to respond to demands for accuracy simply by tinkering with the ways they produce indicators, without bringing outsiders into this process.

On the one hand, ratcheting up formal channels of accountability might channel and manage the demands outsiders make upon an indicator generator. After all, outsiders should usually prefer to avail themselves of institutionalized channels for making demands. Indicator generators might prefer such an arrangement because it lessens the probability that demands will be publicly expressed and hence discredit the indicator.

However, increasing the scope for participation and review also has the potential to markedly increase the volume of demands flowing through the established channels; reducing the cost of making demands should tend to increase their frequency. Moreover, once outsiders have been granted an entitlement to get involved in the generation of indicators, perceptions that participation is not full enough, or that review is not truly impartial or effective, may lead to amplified complaints made outside of these channels. Total exclusion from decision-making can actually be an effective way of discouraging outsiders from attempting to exercise oversight over decisions.

The other accountability components—transparency, reason-giving, and legality—may all tend to increase the stridency of criticisms made of indicator generators, as well. The more outsiders know about how indicators are created, the more likely they are to find fault with these processes. The greater the extent to which generators commit to rules that govern how they develop indicators, the more outsiders can point to procedural shortcomings in order to protest the levels at which indicators are set. In the following section, we will see some reasons why indicator generators, despite these risks, might nonetheless choose to increase levels of transparency and reason-giving. However, it certainly seems unlikely that generators will offer improvements in accountability merely because outsiders demand greater accuracy.

The Supply Hypothesis of accountability

Based on the above discussion, the Demand Hypothesis is likely incorrect in presuming that just because important decisions are based on such indicators, users and targets will attempt to secure enhanced accountability from indicator generators and indicator generators will furnish such accountability. Nonetheless, we should still expect users, when deciding which indicator to employ, to seek the most accurate indicators, all else being equal. This expectation forms the foundation of the Supply Hypothesis.

Just as legislators, in attempting to exercise oversight over development agencies, will usually have difficulty identifying whether development outcomes are a consequence of misallocation of resources by agencies, users will also usually be unable to form *direct* perceptions of an indicator's accuracy. Users should thus tend to judge indicator accuracy not by looking at the indicators themselves, but by looking for

markers that accompany accurate measures. That is, users will usually judge accuracy by evaluating the processes generators use to develop indicators.

If an indicator generator wants to attract as many users as possible, then, the generator should provide those types of accountability that tend to convince potential users that its indicator is accurate. In particular, generators should provide heightened levels of transparency and reason-giving, so that users may be impressed by the thoroughness of a generator's indicator-generating processes and the validity of its methodology. Under this Supply Hypothesis, then, high levels of accountability result not from demands by users or targets, but from competition by indicators to secure the patronage of users. Consequently, the more consultative facets of accountability—participation and review—will tend to be underprovided. Moreover, indicator generators should provide greater transparency and reason-giving when potential users could do without a particular indicator, and when use of an indicator confers a substantial benefit—whether in the form of revenues or influence—upon the indicator generator.

Evaluating the Supply Hypothesis

To test the Supply Hypothesis with reference to our three test cases, we will begin by attempting to ascertain the returns each indicator generator derives from the patronage of its users, and the competitiveness of the environment in which each indicator exists. We will then use the Supply Hypothesis to predict resulting levels of accountability, and compare these levels to those observed in reality.

The user-dependence and competitiveness of the test cases

Both the PRS Group and Freedom House appear to derive significant benefits from the use of these indicators and face competitive markets for their indicators, while the MAR Project enjoys a more secure position in both of these respects.

International country risk guides, PRS Group

The PRS Group appears to depend on revenue from the sale of licenses to its products.¹¹⁵ Because the main purchasers of these licenses are private investors, the Group can be said to rely substantially on its relationship with these investors. The Group also charges academics for the use of the ICRGs, deriving financial benefit from this use.

In contrast, a decision by the MCC to stop using the ICRGs (or the WGI into which they are incorporated) would not directly affect the Group's bottom line,

¹¹⁵ See, e.g., The PRS Group, *International Country Risk Guide: The Rating System and the Publication 2*, available at <<http://www.prsgroup.com/PDFS/ICRG%20BROCHURE-E11.pdf>> (accessed August 5, 2011) (advertising print and online subscriptions to the ICRGs).

since the MCC's use of the ICRGs does not lead directly to increased revenues for the Group. But academics and the MCC confer another sort of benefit upon the PRS Group: when they choose to use the ICRGs in their analyses or decision-making, this results in a higher profile for the Risk Guides and a greater reputation for reliability, both of which should allow the Group to charge higher rates and to sell more licenses to use the Guides. Thus, the PRS Group's dependence on all of its users is significant.

Alternatives to the ICRGs are certainly available¹¹⁶ and internal analysis of governance and political risk is always an option.¹¹⁷ Thus, private investors, academics, and development agencies have other governance measures that they could choose in case the ICRGs appear insufficiently attractive.

Freedom in the World rankings, Freedom House

In contrast to the PRS Group, Freedom House is primarily supported by grants from "various private foundations and government agencies."¹¹⁸ Given that the US government supplies 75 percent of the Freedom House budget,¹¹⁹ it might be true that the organization would see its funding affected if its ratings proved to be of

¹¹⁶ Many governance indicators exist; in her 2003 survey, Marie Besançon lists almost 50. Marie Besançon, "Good Governance Rankings: The Art of Measurement," World Peace Foundation Report No. 36, (2003), 11–34, available at <<http://belfercenter.ksg.harvard.edu/files/wpf36governance.pdf>> (accessed August 5, 2011). Some of these indicators measure only particular aspects of governance (such as Transparency International's Corruption Perceptions Index) or cover only particular regions (such as Asian Barometer), while others are more comprehensive. Transparency International, Global Corruption Barometer (2009), available at <<http://www.transparency.org/content/download/43788/701097/>> (accessed August 5, 2011); Asian Barometer, *Program Overview*, <<http://www.asianbarometer.org/newenglish/introduction/ProgramOverview.htm>> (accessed August 5, 2011) (describing the Barometer as surveying "attitudes and orientations toward political regime, democracy, governance, and economic reform" in East Asia). There is at least one other privately produced political risk indicator—the Business Environment Risk Intelligence (BERI) ratings: Business Environment Risk Intelligence, Business Risk Service, <<http://www.beri.com/brs.asp>> (accessed August 5, 2011) (advertising risk assessment services, including a Political Risk Index). Several other indicators measure the quality of the business environment. See, e.g., The Heritage Foundation, *2011 Index of Economic Freedom*, <<http://www.heritage.org/Index/>> (accessed August 5, 2011).

¹¹⁷ The World Bank uses internal assessments to guide their Performance Based Allocation of funds. See World Bank, "Counter Policy and Institutional Assessment: Frequently Asked Questions," (2010), available at <<http://www1.worldbank.org/operations/CPIA2010/webFAQ10b.pdf>> (accessed August 5, 2011). The Asian Development Bank, see Asian Development Bank, *Refining the Performance-Based Allocation of Asian Development Fund Resources* (October 2008), available at <http://www.adb.org/Documents/Policies/ADF/Performance_Based_Allocation/Refining-Performance-Based-Allocation.pdf> (accessed August 5, 2011), and the African Development Bank also follow this model. African Development Bank Group, *2010 CPIA Questionnaire* (October 2010), available at <<http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/2010%20CPIA%20Questionnaire%20-%20July%202010%20EN.pdf>> (accessed August 5, 2011). As we have already discussed, moreover, many investors may base their decisions mostly on their own assessments, using services like the ICRGs only in supplementary fashion.

¹¹⁸ Freedom House, About Us, <<http://www.freedomhouse.org/about-us>> (accessed April 12, 2012).

¹¹⁹ See Freedom House, *Annual Report* (2005), 24 (stating that of \$26.8 million in total revenues and support received in Fiscal Year 2005, \$20.2 million came from federal grants).

little use to US development agencies like the MCC. However, given that Freedom House received US funding for many years before the MCC started using its ratings to guide funding allocation, and given the reasons we have already elucidated for questioning the interest of officials in aid efficacy, this connection seems tenuous and unlikely. On the other hand, Freedom House is an advocacy NGO that likely measures its success and justifies its existence to funders based in part on how influential its ratings are. Consequently, decisions by journalists, academics, or development agencies to use other indicators as measures of good governance would certainly injure the organization's interests.

As with the ICRGs, moreover, the Freedom in the World ratings have many competitors, though some of these indicators may not measure the same phenomena as Freedom House, cover the same broad range of countries, extend as far back in history, or possess Freedom House's high profile. Thus, some users may default to use of Freedom House's ratings even if they have little reason to believe that the ratings are reliable, basing their choice not on probable accuracy but on reputation or the need for extensive data. We will return to this point when we consider the defects of the Supply Hypothesis.

Minorities at Risk variables, MAR Project

Finally, the MAR Project does not derive financial support from its users; the MAR variables are free.¹²⁰ On the other hand, though, the users of the MAR indicators may provide much in the way of influence, visibility, and usefulness to the indicators, the quest for which presumably supplies at least some of the motivation for the indicators' generators. Another important aim of the MAR Project and its funders might be simply to expand the amount of available knowledge about ethnopoltical groups, with only limited concern about how this knowledge is used. We might expect, then, that the MAR Project will be less concerned with expanding its field of users than the PRS Group or Freedom House. As for levels of competition, the MAR Project is quite unique in studying and coding the mobilization and treatment of ethnic and political groups; researchers interested in studying these issues have few alternatives to which they can turn.

Comparing predictions to observations

We would predict, on the basis of this analysis, that the PRS Group should have the highest levels of transparency and reason-giving, while Freedom House would have perhaps slightly lower levels, and the MAR Project the lowest levels of all. Levels of participation and review should be low throughout all the cases.

Are these predictions borne out by reality? As discussed above in relation to the Demand Hypothesis, PRS Group and Freedom House both provide significant

¹²⁰ See MAR (n. 12), Data.

transparency and reason-giving and the MAR Project does not, though in contrast to predictions, Freedom House actually surpasses the PRS Group in the supply of these components. Neither the PRS Group nor Freedom House offers much room for participation or review, though the MAR Project does provide for substantial participation. The Supply Hypothesis, then, cannot be said to represent the whole story of accountability in indicator generation, but it seems to have more explanatory power than the Demand Hypothesis. By considering where this Hypothesis falls short, we can construct a more complete explanation.

Explaining the failures of the Supply Hypothesis

The Supply Hypothesis likely suffers in large part from excessive simplicity. In particular, the Hypothesis envisions no role for reputation or identity.

The role of reputation

A user may judge the accuracy and reliability of an indicator from the methodology by which it is set and the reasons given to justify indicator levels, but these qualities may also be judged by reference to the indicator's reputation. If an indicator is widely used and little criticized, would-be users can be more confident that it is reliable, and the indicator generator will have to expend less energy convincing users of this fact through enhanced transparency and reason-giving. If an indicator is obscure or frequently pilloried, however, generators may have to work harder to appear reliable.

The role of reputation in bolstering perceptions of accuracy may explain why the ICRGs are supported by less transparency and reason-giving than the Freedom in the World ratings. The Freedom House ratings, while well-known, are also often subject to charges from journalists and academics that they display a right-wing bias and are influenced by the political aims of the US government.¹²¹ The ICRGs, though widely used, are rarely the object of such critical coverage. The PRS Group may thus be more secure in relying on its reputation to attract users than Freedom House.

Identity and participation

The Supply Hypothesis predicts that generators will not make provision for participation or review in order to "sell" their indicators to prospective users. In

¹²¹ See, e.g., Right Web, Freedom House, <http://www.rightweb.irc-online.org/profile/Freedom_House> (accessed August 5, 2011) (noting that "[a]lthough in recent years the organization has appeared to relax its close association with hawkish U.S. policies, its leadership remains heavily represented by individuals affiliated with neoconservatism and it has continued to support projects aimed at bolstering aggressive U.S. foreign policies"); Bollen, "Political Rights and Political Liberties in Nations" (n. 7), 205 (noting that "researchers claim that [Freedom House's] conservative slant has led to incorrect rights and liberties ratings for some countries").

our discussion of the defects in the Demand Hypothesis, moreover, we reasoned that generators should be disinclined to furnish high levels of participation or review under most circumstances, since ratcheting up levels of these accountability components might tend to prompt more demands and criticisms from outsiders. These two propositions nonetheless leave room for generators to allow participation and review under some circumstances. When (1) it is not too costly to bring outsiders into the indicator development process; and (2) the outsider-generator relationship is marked by a certain level of trust, generators may increase the levels of participation and review that are available. In short, when indicator generators and a certain type of outside actor share a single community, provision of participation and review should rise.

The best example of this principle from our case studies arises in the context of the MAR Project. Participation by users in that case seems to have been enabled by the fact that MAR users and indicator generators shared a profession, so that they interacted at professional conferences and felt comfortable exchanging information on indicator reliability. The MAR Project generators did not have to worry about demands from their academic users spiraling out of control, because the generator-user relationship was likely marked by collegiality and civility. The Project also did not have to invest many resources in setting up channels of participation. In fact, shared identity can explain more in the MAR context than just the observed levels of participation. The shared profession of MAR generators and users may also explain whatever transparency deficits we observe, as well the absence of formal review. When academic users of the MAR indicators need additional information about how variables are coded, they likely can simply contact the Project administrators for additional information; they may be able to question the levels at which indicators have been set through similar informal channels, without the need for recourse to an independent arbitrator.¹²²

Conclusion

This chapter has sought to explain what determines levels of accountability in indicator generation. It first considered a Demand Hypothesis, under which indicator users and targets made demands for improved accuracy or accountability

¹²² In an earlier footnote, we suggested that secretively elevated levels of accountability for some users undermines accountability in general. See above nn. 56, 106. Selectively elevated levels of accountability may appear similarly harmful, but they do not undermine legality as much as secretively elevated levels; as long as it is evident that some actors have more access to the indicator development process, other actors can adjust their demands and expectations accordingly. Even if the MAR Project does furnish its users with expanded accountability in some non-transparent, irregular ways, it matters whether this apparent secretiveness is the result of an intent to conceal demands and complaints being made, or whether it is merely a function of the fact that few outside the community of users really care about the indicator. If the only real audience for an indicator is its users, and these users share an identity with the indicator generator and hence enjoy higher levels of accountability, we should not consider the overall accountability profile of the indicator to have been injured. The generator is still providing for voice by those who would exercise it.

when consequential decisions were based on indicators. When this hypothesis failed to describe the test cases, the analysis then turned to a Supply Hypothesis, under which competition for users prompted indicator generators to provide certain kinds of accountability. This hypothesis encountered more success, though it too appeared to have flaws.

A synthesis of this analysis suggests that indicator users and targets will rarely demand either accuracy or accountability from generators, even when important decisions are based on indicators. Indicator generators should display higher levels of transparency and reason-giving when they are in competition with other indicators or users can do without indicators altogether, and when users provide the generators with significant revenue or influence. However, these levels should be moderated when a generator has already established a reputation for reliability. High levels of participation and review should generally not be observed in indicator generation, though these accountability components might be furnished when generators and users (or, conceivably, targets or other kinds of actors) share an identity attribute that decreases the costs of furnishing such accountability for both types of actors.

Of course, these conclusions are based on the tentative study of only three governance indicators; future research examining accountability in the generation of other indicators would be most useful. Furthermore, these conclusions may be generalized rather easily to other instances of global administration. Whether they hold in such contexts might be a promising avenue for further investigation.

Public Regulation of Global Indicators

*Sabino Cassese and Lorenzo Casini**

Introduction: Indicators in global governance and their regulation

Indicators are today critical to governance, national and global, in part because they consolidate power in the hands of those with a particular kind of “knowledge.”¹ The massive use of data and statistics by national governments and international organizations triggers the development of indicators amongst public administrations working in many different sectors, including health, education, and the environment.² Significant institutional examples include the World Bank, the World Health Organization, and the United Nations Development Programme. International finance has itself become saturated with indicators.³ And indicators increasingly permeate the development and administration of human rights policy.⁴

* The authors warmly thank Eleanora Cavalieri, Sarah Dadush, Maurizia De Bellis, Angelina Fisher, Euan Mac Donald, and Giulio Napolitano for their helpful comments.

¹ An overview is in Kevin E. Davis, Benedict Kingsbury, and Sally Engle Merry, “Indicators as a Technology of Global Governance,” *Law and Society Review* 46:1 (2012): 71–104, who underline the increasing production and use of indicators in global governance. According to these authors, “An indicator is a named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis (such as countries or institutions or corporations), synchronically or over time, and to evaluate their performance by reference to one or more standards.” See also Christiane Arndt and Charles Oman, *Uses and Abuses of Governance Indicators* (Paris: OECD Development, 2006).

² Public authorities have always kept control over the management of weights and measures. They have been measuring and numbering for centuries: censuses, for instance, were carried out as early as the second millennium BC, such as in the famous biblical precedent when God ordered Moses to take a census of the entire Israelite community (Numbers, 1:1–3). Nowadays, every state has its own statistics office or division that collects data, conducts research, and produces reports and surveys. Take, for example, FedStats in the United States. Similar agencies also operate at the supranational level: this is the case of Eurostat within the EU—which was established in 1953 to meet the requirements of the Coal and Steel Community, even before the foundation of the European Economic Community—or the United Nations Statistics Division (UNSD) of the Department of Economic and Social Affairs (DESA).

³ See Terence C. Halliday, “Legal Yardsticks: International Financial Institutions as Diagnosticians and Designers of the Laws of Nations,” this volume, pp. 180–216.

⁴ On the increasing use of indicators in the field of human rights, see Margaret L. Satterthwaite and AnnJanette Rosga, “Measuring Human Rights: UN Indicators in Critical Perspective,” this volume, pp. 297–316 and Ronen Shamir and Dana Weiss, “Semiotics of Indicators: The Case of Corporate Human Rights Responsibility,” this volume, pp. 110–31.

Indicators are spreading globally because globalization has triggered, among other legal effects, a need to compare economic performance of states and firms, and populations' social conditions in order to rate and to rank legal systems based on, for example, their competitiveness or their degree of "freedom." Comparisons are essential to better understand world phenomena, and indicators are a convenient way to penetrate different legal, social, and economic systems.⁵ Moreover, indicators play a big role in levelling the "playing field" in which states and supranational actors operate, because they can promote imitation and transplants between legal systems and help reduce the fragmentation produced by the formation of several global regulatory regimes. In addition, global governance requires balancing uniformity and diversity; indicators measure diversity, often attempting to attach a value to it, but set as a measurement reference point certain (seemingly universal) criteria, the selection of which can be contentious.⁶ Finally, the "law and development" movement has favored the use of indicators, which emerged in the 1960s and 1970s, and has grown significantly over the past decades.⁷ Within this movement, the ability to quantify and compare legal and social systems is considered not only informative, but also conducive to the achievement of development goals.⁸

One result of the expansion of indicators globally is that a growing number of important decisions rely on them. Particularly where the decision-makers are international organizations dealing with the supply and safeguarding of public goods and rights, the question arises how to regulate the production and use of these tools, triggering other questions regarding accountability and legitimacy.⁹ Indicators considered "per se" regulatory devices may be intentional or unintentional.¹⁰ This is why some commentators argue that legitimacy of global indicators can be enhanced only by developing a robust regulatory framework to oversee their production and use.¹¹

This chapter addresses these issues by analyzing the process of regulating the production and use of global indicators. In particular, we claim that while indicators *can* display an intrinsic "regulatory" nature, not all require insertion into

⁵ See Tim Büthe, "Beyond Supply and Demand: A Political-Economic Conceptual Model," this volume, pp. 29–51.

⁶ Take, for example, the Country Policy and Institutional Assessment (CPIA) developed by the International Development Association (IDA)—i.e., the institution of the World Bank responsible for providing financing to the poorest countries—which evaluates countries on the basis of 16 criteria (e.g., economic management and public sector management and institutions: Davis, Kingsbury, and Merry, "Indicators As a Technology of Global Governance" (n. 1), 92.

⁷ See John H. Merryman, "Comparative Law and Social Change: On the Origins, Style, Decline & Revival of the Law and Development Movement," *The American Journal of Comparative Law* 25 (1977): 457.

⁸ A significant case is the World Development Indicators annually produced by the World Bank. Some commentators, however, observed that the use of indicators in this setting generates controversy: see, for instance, Andrew Natsios, "The Class of the Counter-bureaucracy and Development," a Center for Global Development Essay, (July 2010) at <<http://www.cgdev.org/content/publications/detail/1424271>> (hereafter "Natsios").

⁹ See Nikhil K. Dutta, "Accountability in the Generation of Governance Indicators," this volume, pp. 437–64.

¹⁰ Wendy Espeland and Michael Sauder, "The Dynamism of Indicators," this volume, pp. 86–109.

¹¹ Armin von Bogdandy and Matthias Goldmann, "Taming and Framing Indicators: A Legal Reconstruction of the OECD's Programme for International Student Assessment (PISA)," this volume, pp. 52–85.

regulatory frameworks in order to ensure their legitimacy and accountability. Whether a framework is needed depends on several factors, such as the type of indicator at issue, characteristics of the indicator-producing entity, and the kinds of users involved. Thus this chapter purports to set out a taxonomy of different types of indicators, in order to distinguish cases in which indicators serve as tools for holding governments accountable from those in which indicators, as accountability-enhancers, themselves require oversight by a legitimate authority.

Regulatory implications of global indicators

As mentioned above, indicators can have an intrinsic regulatory nature. This raises significant questions, such as: “How does an indicator become a legal device?” and “Which indicators are legitimate sources of power?” Finally, “If indicators are not only pieces of information, but also technologies of global governance, should they be subject to external controls or oversight?”

From scientific “authority” to regulatory “authority”: A taxonomy

In the first instance, indicators are measurements. As such, they can derive scientific “authority” from their accuracy and from the legitimacy of their producer. But indicators can also have regulatory “authority” that is derived perhaps less from their design than from their use. This type of authority can be difficult to ascertain because indicators are often multipurpose devices, serving different functions for different users. Varying production processes and uses allow the identification of two main categories of indicators: *binding* and *voluntary*.¹²

Binding indicators are those imposed on or applied to “regulatees” by states, international governmental organizations, or other global institutions (either public or private, especially those which administer funds or allocate capital or other resources). These indicators help regulators evaluate, assess, and make decisions. While the regulatees or addressees do not choose the indicators, they are often compelled to adapt their behavior or operations in order to demonstrate good performance or compliance with the standards disseminated through such indicators.

Furthermore, within this category, two different types of indicators exist. First, there are indicators produced by regulators and imposed on or applied to regulatees. The term “regulator” here includes not just governments, but any other different bodies deploying indicators. Take, for example, the indicators of the Global Fund to Fight AIDS, Tuberculosis, and Malaria: these are used by the Fund in order to

¹² The taxonomy here proposed strictly refers to different geneses and uses of indicators. It would, of course, require further empirical investigation, and it does not exclude other possible categories; for example, those based on the legal regime under which indicators are produced (public or private indicators), as well as on the kind of promulgator (formal inter-governmental organization, inter-governmental network, hybrid public and private body, private entity, state) or on the type of indicators (index or ranking).

determine the allocation of financing to civil society organizations involved in the fight against the diseases.¹³

Secondly, indicators can be used by third parties (not the producer or promulgator, nor the regulatee or the party addressed by the indicators) to evaluate the regulatees. A useful example is the set of indicators used by the Programme for International Students Assessment (PISA), which concerns student evaluation: here, third parties, namely national authorities, adopt the indicators produced and regularly updated by the Organization for Economic Co-operation and Development (OECD), thereby conferring authority upon them.¹⁴ Other examples are the "Doing Business" set of indicators produced by the World Bank, but used by all kinds of third parties, private and public,¹⁵ or ratings generated by specialized agencies in the credit sector.¹⁶

Rather than being imposed by an authority, voluntary indicators are chosen. For this category, we can again distinguish two hypotheses. First, indicators can be used by the regulatees themselves, in a process of voluntary compliance, as a means of self-coordination. For example, the Human Development Indicators are instrumental in pressuring slow performing countries to conform to the standards contained within and disseminated by the indicators, and to improve performance accordingly.¹⁷ It should be noted that compliance through conformity is one of the characteristic features of global administrative law,¹⁸ and that, from this perspective, the use of voluntary indicators illustrates the spread of compliance mechanisms based on factors other than enforcement, such as imitation or competitiveness.

Secondly, indicators can be used by third parties to influence other third parties. This happens, for example, when activist groups use the US State Department Trafficking in Persons indicators to influence prospective tourists,¹⁹ or when media

¹³ The production and use of indicators in public health at the global level is analyzed by Angelina Fisher, "From Diagnosing Under-immunization to Evaluating Health Care Systems: Immunization Coverage Indicators as a Technology of Global Governance," this volume, pp. 217–46.

¹⁴ See von Bogdandy and Goldmann, "Taming and Framing Indicators" (n. 11).

¹⁵ See Kevin E. Davis and Michael B. Kruse, "Taking the Measure of Law: The Case of the Doing Business Project," *Law & Social Inquiry* 32 (2007): 1095.

¹⁶ See Frank Partnoy, "The Paradox Of Credit Ratings," University of San Diego School of Law, Law and economic research paper, no. 20, 2001, and "How And Why Credit Rating Agencies Are Not Like Other Gatekeepers," University of San Diego School of Law, Research Paper no. 07–46, May 2006; Claire A. Hill, "Regulating The Rating Agencies," *Washington University Law Quarterly* 82 (2004): 43, and Steven L. Schwarcz, "Private Ordering Of Public Markets: The Rating Agency Paradox," *University of Illinois Law Review* 1 (2002).

¹⁷ On the Human Development Index, see Davis, Kingsbury, and Merry, "Indicators As a Technology of Global Governance" (n. 1), 95–9.

¹⁸ See Sabino Cassese, "Is There a Global Administrative Law?" in Armin von Bogdandy et al. (eds), *The Exercise of Public Authority by International Institutions. Advancing International Institutional Law* (Heidelberg: Springer, 2010), 772ff; see also Benedict Kingsbury, "The Concept of 'Law' in Global Administrative Law," *European Journal of International Law* 20 (2009): 23, and David Dyzenhaus, "Accountability and the Concept of (Global) Administrative Law," in *Global Administrative Law* (Acta Juridica: Cape Town, 2009), 3ff.

¹⁹ On these indicators, see Anne T. Gallagher and Janie Chuang, "The Use of Indicators to Measure Government Response to Human Trafficking," this volume, pp. 317–43 and Marina Zaloznaya and John Hagan, "Fighting Human Trafficking or Instituting Totalitarian Control? The Political Co-optation of Human Rights Protection in Belarus," this volume, pp. 344–64.

or NGOs refer to the ratings produced by Freedom House.²⁰ In this case, indicators become instruments of governance in so far as they can stimulate or provoke significant policy interventions in countries that fail to achieve a high ranking.

Finally, indicators can be used in more than one of these modes.²¹ On the one hand, regulators themselves develop different uses of the indicators they generate. The World Bank, for instance, produces indicators that are used both by the Bank itself and by other institutions in deciding how to allocate aid. This transposition or appropriation of indicators by other users can happen either for a different purpose or just for the same use, expanded.²² On the other hand, indicators can be considered as being simultaneously binding and voluntary, depending on several factors, including their scope, their addressees, and their producers. At the same time, some indicators could first be produced as voluntary tools but then turn into binding ones; conversely, the latter may “lose” their authoritative character and be progressively overtaken or rendered irrelevant by other indicators. In other words, the taxonomy provided here should not be considered rigid; the categories tend to be fluid and flexible, depending on the different histories, contexts, and dissemination methods attached to different indicators.

Legitimacy and accuracy

The taxonomy of different geneses and uses of indicators helps provide a framework for examining other questions related to indicators’ legitimacy, regulation, and control.

In order to better understand the issue of legitimacy, it is worth recalling that indicator *uses*, rather than the indicators themselves, are relevant.²³ It is when a regulator produces and publishes indicators, and takes decisions on the basis of those indicators, that the question of legitimacy arises. In that setting, it is crucial to know what type of authority was conferred on the producer and promulgator of a given indicator, whether there are any conflicts of interest, what procedures were followed in the production process, on the basis of what data the indicator was produced, and finally, which body or bodies can review the indicator. In considering the legitimacy of binding indicators, two main hypotheses can be distinguished.

First, indicators can be the final product of activities carried out by public administrations (e.g., statistics offices or divisions) or by international governmental organizations. Under those circumstances, the legitimacy of indicators is similar

²⁰ See <<http://www.freedomhouse.org/template.cfm?page=1>>.

²¹ Governing with indicators is not a one-way, top-down process, and therefore we need a more robust theory of global influence and authority. This point has been highlighted by René Uruña, “Internally Displaced Population in Colombia: A Case Study on the Domestic Aspects of Indicators as Technologies of Global Governance,” this volume, pp. 249–80. Christopher Stone’s “Problems of Power in the Design of Indicators of Safety and Justice in the Global South,” this volume, pp. 281–94, also considers a bottom-up approach in producing global indicators.

²² Davis, Kingsbury, and Merry, “Indicators As a Technology of Global Governance” (n. 1), at 84.

²³ This point is also underlined by Katharina Pistor in “Re-construction of Private Indicators for Public Purposes,” this volume, pp. 165–79.

to that accorded to any other public function. Therefore, when indicators are produced by regulators and imposed on or applied to regulatees, they simply represent another kind of regulatory intervention. As a consequence, indicators have the same legitimacy as other manifestations of public functions; specifically, they are as legitimate as the institution that produced them.

Secondly, indicators can be used by parties other than the producer or promulgator in order to evaluate regulatees (as with the PISA indicators). Here, the legitimacy of the use of the indicator cannot be derived solely from that of its producer, which can raise an array of additional problems. On what basis do states or global institutions choose one particular indicator rather than another? How can regulatees be protected against decisions taken by regulators in accordance with data and criteria produced by other regulatory bodies and other procedures (i.e., the procedures in which those indicators have been produced)?

Turning to voluntary indicators, the situation seems to be simpler. If indicators find their legitimacy through a bottom-up mechanism—i.e., they are freely chosen by the regulatees, or by third parties—then the legitimacy concern should not arise. Indeed, with voluntary compliance, the use of specific indicators is not imposed; rather, it is based on consensus. While this distinction is useful for sketching a preliminary taxonomy, it is crucial to bear in mind that case-by-case analyses are needed to verify whether voluntary indicators are truly chosen by regulatees or are instead somehow imposed through other mechanisms. On the other hand, “voluntary” indicators might in fact be sustained by non-regulatory forms of accountability and compliance, based on market or reputational pressures, for instance, which may in turn strengthen the legitimacy attached to the consensus.²⁴

The test of repeatability (or test-retest reliability) can be extremely effective in establishing whether there is a problem of legitimacy or a problem of accuracy.²⁵ If the same value (e.g., human development, democracy, transparency) can be gauged with the same measurements or with different indicators produced by a different institution—or, in other words, if there are multiple indicators available to assess the same phenomenon—then it could be argued that there is less need for regulation because the indicators will compete amongst themselves to produce the most accurate results and as a consequence, to enhance their respective legitimacy.

As an example of this phenomenon, consider the case of the World Bank’s “Doing Business” indicators. These are founded on the presumptions that less regulation is better than more regulation, and that common law is better than civil law for ensuring economic growth. This triggered a French reaction²⁶ that led to

²⁴ Ruth W. Grant and Robert O. Keohane, “Accountability and Abuses of Power in World Politics,” *American Political Science Review* 99 (2005): 29, 37.

²⁵ This method was first developed in clinical measurement by J. Martin Bland and Douglas G. Altman, “Statistical Methods For Assessing Agreement Between Two Methods of Clinical Measurement,” *The Lancet* (February 8, 1986): 307. That study highlighted that “repeatabilities” of methods of measurement “limit the amount of agreement which is possible.”

²⁶ See, for instance, Bertrand du Marais, Didier Blanchet, and Anna Dorbec, *Des Indicateurs Pour Mesurer Le Droit? Les Limites Methodologiques Des Rapports Doing Business* (Paris: La documentation française, 2006).

the establishment, in May 2007, of the “*Fondation pour le Droit Continental*—Civil Law Initiative,” which seeks to contribute to and enhance the eminence and international influence of the continental legal and judicial tradition, and of French law, especially in the field of commercial law, and the wider law relating to business and the economy; the influence and strategic position of French law within international bodies, in relation to public and private-sector decision-makers; the international profile of legal agencies; and the development of legal research, with emphasis on scientific comparative studies, especially studies designed (in the interests of economic development) to increase the attractiveness of the French legal system.²⁷

As this example illustrates, discussions pertaining to indicator-adequacy can take place outside the context of the promulgator, reaching into civil society, particularly where the social consequences of an indicator’s application are significant. Perhaps the French foundation will develop a different and more accurate set of indicators, without, however, identifying as its exclusive purpose the toppling of the “Doing Business” indicators.

Bringing indicator production under control

Lastly, if indicators represent a technology of global governance, should they be subjected to regulation?

Approaches to the problem of regulating indicators are very diverse. This variety is illustrated by the case of credit rating agencies. Under one approach, ratings are protected by free speech. Therefore, indicators produced by credit rating agencies are and should remain free from regulation. This is what happened in the United States until recently.²⁸ Due to the growing importance of credit agencies, however, a second approach emerged that does not create additional oversight over the production of indicators or subject their promulgation to regulation, but does subject the *producer and promulgator* to legislation. This is what happened in the United States with the 2006 Credit Rating Agency Reform Act, and also in the European Union. The 2009 European regulation on credit agencies (Reg. 1060/2009) provides norms regarding registration, conflict of interest, disclosure, surveillance, and sanctions at the European and at the national level.²⁹ A separate regulation (Reg. 1095/2010) established a European body responsible for oversight of credit rating agencies (the European Securities and Markets Authority—ESMA).³⁰

²⁷ At <http://www.fondation-droitcontinental.org/jcms/c_5105/accueil>.

²⁸ Monica R. Das Gupta, “The External Accountability Gap Of Private Regulators: Accountability Paradoxes And Mitigation Strategies. The Case Of Credit Rating Agencies,” *International Public Policy Review* 1 (2005): 37.

²⁹ In 2011, EU Reg. No. 513/2011 amended Reg. No. 1060/2009 in order to endow ESMA with more significant powers over credit rating agencies, namely regarding registration, surveillance, and sanctions.

³⁰ Fabian Amtenbrink and Jakob De Hann, “Regulating Credit Ratings in the European Union: A Critical First Assessment of Regulation 1060/2009 on Credit Rating Agencies,” *Common Market Law Review* (2009): 1915.

A third approach requires that neither the producer, nor the indicators be subjected to regulation, but that the *process* by which indicators are produced and promulgated be regulated. Here, administrative law principles—such as transparency, access, participation, and review—can be very helpful.³¹ In the public health sector, for instance, the WHO is currently assessing the possibility of introducing more structured forms of participation for NGOs, corporations, and civil society actors in its regulatory activities.³² In other words, this approach seeks to improve the accountability of indicators through procedural mechanisms and principles, such as notice and comment, hearings, or committees.

Finally, there is the view according to which indicators themselves must be subjected to regulation or produced by public authorities. This argument has been made with regard to certain private indicators³³ and it is a common trend in global private regimes: the more indicators develop and gain prominence, the more public authorities attempt to regulate and limit them.

It is not pre-determined which of these approaches is best. Once again, it is helpful to consider how a given indicator is used, and by whom, in order to determine which regulatory model is most appropriate.

In the case of voluntary indicators, where compliance is based on consensus, there may not be a need for regulation. However, there can be cases in which third parties create pressure on states or international organizations to adopt the standards contained in and disseminated through indicators. In fact, indicators and their private promulgators might be at their most dangerous when they stand in for government, because in those instances, nominally voluntary compliance quickly becomes involuntary.

With regard to binding indicators, when it is the regulator itself which produces the indicators, there is no need for special regulation. As mentioned above, indicators will follow the rules set forth for a specific field of activity; and, in these cases, rulemaking procedures should of course be transparent, especially with respect to the criteria underlying or supporting the indicators at issue. If, however, an indicator that is produced and promulgated by one institution is then used by another institution in evaluating a third party, then the need for a legal framework emerges. This is the case, for instance, with the student evaluations carried out by the OECD. This gives rise to a new question: should the regulatory framework

³¹ Fisher, "From Diagnosing Under-immunization to Evaluating Health Care Systems" (n. 13).

³² See Gaudenz Silberschmidt et al., "Creating a Committee C of the World Health Assembly," *The Lancet* 371 (May 3, 2008): 1483ff. In fact, WHO relies on committees of experts, largely appointed by the Organization itself, operating in private without wide participation of states or corporations. However, much of the pressure by states or non-state actors (including pressure from the EU) for more participation aims to give voice to special interests, in some cases those seeking direct commercial advantage, leading to calls for the WHO to continue to resist. See Benedict Kingsbury and Lorenzo Casini, "Global Administrative Law Dimensions of International Organizations Law," in *Symposium on Global Administrative Law in the Operations of International Organizations* (ed. Laurence Boisson de Chazournes, Lorenzo Casini, and Benedict Kingsbury), *International Organizations Law Review* 6 (2009): 319, 355.

³³ See Sarah Dadush, "Impact Investment Indicators: A Critical Assessment," this volume, pp. 392–434.

apply to the OECD or to the national authorities using OECD indicators for regulatory purposes?³⁴

These approaches are not mutually exclusive; they are often adopted together, as happened in the case of credit rating agencies. This example again helps show how a global regulatory system for indicators could be established.

In particular, we argue here that the emergence of global regulation resembles the model built in the 1950s by Oliver MacDonagh in order to explain the genesis of the “governmental revolution” in Britain during the middle of the nineteenth century.³⁵ Based on the “learning by experience” approach, that model has five stages. The *first* is triggered by exposing a severe social harm (e.g., child labor, mining accidents), followed by demands for a remedy, and the subsequent adoption of legislation.³⁶ In the *second* stage, new findings of harms are made public, revealing that the original ones have not been remedied by the adopted legislation because it lacks an enforcement mechanism. These discoveries lead to the *third* stage, where central administrators are appointed who acquire expertise in their respective fields and become a vocal force in favor of additional legislation and greater centralization. In the *fourth* stage these centralized bureaucrats view their mission as a permanent one, replacing a static concept of administration with a dynamic one, wherein they play a leading role in closing loopholes and “tightening the screws.”³⁷ Finally, in the last stage, these bureaucrats become convinced by the idea of scientific expertise and adopt a more or less conscious Fabianism.

The global regulation of credit rating agencies followed a similar path of seven stages, very rapidly, from 2004 to 2010.

- (1) In 2004, the International Organization of Securities Commissions (IOSCO) Code of Conduct (“Fundamentals for Credit Rating Agencies”) was adopted, based on the “comply or explain” principle: obligation to publish ratings methodologies and to separate ancillary services;
- (2) in 2006, the above-mentioned US Credit Rating Agency Reform Act was approved in order to regulate the conflicts of interest in credit rating agencies and the issuing of credit ratings;
- (3) in 2007, IOSCO and the Committee of European Securities Regulators (CESR) published reports on credit rating agencies compliance with the IOSCO Code of Conduct;

³⁴ An in-depth analysis of these issues appears in von Bogdandy and Goldmann, “Taming and Framing Indicators” (n. 11).

³⁵ Oliver MacDonagh, “The Nineteenth-Century Revolution in Government: A Reappraisal,” *Historical Journal* (1958): 56; see also Roy M. MacLeod, “Statesmen Undisguised,” *American Historical Review* (1973): 1386; John H. Goldthorpe, “Le Développement De La Politique Sociale En Angleterre De 1800 à 1914,” *Sociologie du Travail* (1963): 105; Daniel Norman Chester, *The English Administrative System 1780–1870* (Oxford: Clarendon, 1980); and Sabino Cassese, *La Construction du droit administratif: France et Royaume-Uni* (Paris: Montchrestien, 2000).

³⁶ In this first stage, resistance was encountered from the endangered interests, but instead of inaction, the usual result was a weak compromise.

³⁷ MacDonagh, “The Nineteenth-Century Revolution in Government: A Reappraisal” (n. 36).

- (4) in 2008, the Code was extended to rating of derivatives;
- (5) in 2009, the Financial Stability Forum urged credit rating agencies to comply and asked national authorities to monitor implementation; at the same time, the Financial Stability Board and the G20 urged governments to establish registration of credit agencies;
- (6) in 2009, the above-cited European regulation on credit rating agencies (Reg. 1060/2009) was enacted, providing norms regarding registration, conflicts of interest, disclosure, surveillance, and sanctions at the European and at the national level;
- (7) in 2010, the European body with oversight of credit rating agencies (European Securities and Markets Authority—ESMA) was established; and in 2011 its powers regarding registration, surveillance, and sanctions were expanded. In the United States, the Dodd-Frank Act introduces a stricter regulation of credit rating agencies (corporate governance, disclosure, ratings of structured finance products, sanctions) and establishes an Office of Credit Rating within the US Security and Exchange Commission (SEC).

This path displays several similarities with the nineteenth-century British “governmental revolution,” due to the interaction between preliminary remedies, research and reports, ad hoc administrations, growing scientific and technical expertise, and further regulations. Put briefly, the process is largely self-sustaining and self-generating, based on practice, and it is a process in which changes are led by the administrators who are themselves involved in a “learning by experience process.”

Within this framework, indicators represent one of the most promising forms of global regulation: not surprisingly, considering that the word “indicator” has the same etymology as the word “dike,” the name of the ancient Greek goddess of justice.³⁸ Indicators, however, are a technique of indirect rule or of governance at a distance. Not all of them have an intrinsic “regulatory” nature, and there may be no need to subject all indicators to a legal framework or to regulation. It therefore becomes crucial to distinguish between different hypotheses supporting additional oversight based on the geneses and uses of indicators, as well as on their legitimacy and accountability.

³⁸ Both words, in fact, come from Greek “δείκνυμι,” derived from Sanskrit “dicati,” meaning “to show,” “to indicate.”

Index

Introductory Note

References such as “178–9” indicate (not necessarily continuous) discussion of a topic across a range of pages. Wherever possible in the case of topics with many references, these have either been divided into sub-topics or only the most significant discussions of the topic are listed. Because the entire work is about “indicator” and “governance” the use of these terms (and certain others which occur constantly throughout the book) as an entry point has been restricted. Information will be found under the corresponding detailed topics.

- Abbott, A.** 203, 208, 216
Abbott, K.W. 11, 36
absorption uncertainty 165
abstract level 304–5, 424
abstraction 72–3, 139, 153, 211, 316
Academic Ranking of World Universities (ARWU) 87, 100–3, 106–7
academic validation 177–9
accountability 20–1, 291, 301–2, 367, 369–70, 374, 466–8
 components 439–40, 458, 463–4
 demand hypothesis of
 barriers to collective action 456–7
 behavior of indicator targets 454–7
 comparison of predictions and observations 451–2
 evaluation 445–51
 explanation of failures 452–8
 Freedom in the World ratings as test case 442–3
 hedging by users 452–3
 inability to marshal credible threats of exit 455–6
 insulation from consequences of decisions 453
 International Country Risk Guide (ICRG) as test case 440–2
 lack of credibility or means 454–5
 methodologies of test cases 446–8
 Minorities at Risk (MAR) as test case 444–5
 observed levels of accountability 448–51
 predictions 445–6
 reluctance by generators to bolster accountability 457–8
 unwillingness by users to undermine indicators 453–4
democratic 299
in the generation of governance
 indicators 437–64
and impact investment indicators 422–3
levels of 27, 438, 445, 452, 454, 457, 463
observed levels of 448–51
reluctance by generators to bolster 457–8
supply hypothesis of 27–8, 438, 458–64
comparison of predictions and observations 461–2
evaluation 459–62
explanation of failures 462–3
Freedom in the World ratings as test case 460–1
identity and participation 462–3
International Country Risk Guide (ICRG) as test case 459–60
Minorities at Risk (MAR) as test case 461
role of reputation 462
user-dependence and competitiveness of test cases 459–61
accounting matrix 147–8
accuracy 27, 84, 231, 438–40, 452–4, 456–8, 461–2
 enhanced/improved 439, 453–7, 463
ACER *see* Australian Council for Educational Research
active indicators 25, 284–93
Active Learning Network for Accountability and Performance (ALNAP) 370
activism, judicial 274, 277, 279
activists 252–3, 263, 273, 349, 444–5
actor mismatch 212–13
actor-network-theory 5, 86, 111–12, 135
adaptations, behavioral 30, 32–3, 51
ADB *see* Asian Development Bank
administrative data 230, 243, 293
Advanced Market Commitment (AMC) 219, 233–4
Afghanistan 117, 132, 255, 333
African Development Bank (ADB) 196, 460
aggregate indices 168, 172, 174, 177, 179
aggregation 6, 17, 25, 138, 146–8, 150–1, 153–5
agriculture 317, 396–7, 401, 408
aid 15, 35, 150, 236–8, 258–9, 380–2, 453–4
 agencies 21, 141–2, 439
 humanitarian 268, 372, 385
Albania 16, 58, 187–8, 323, 333
alignment 53, 227, 289–90, 294
Alkhovka, I. 355–7

- ALNAP *see* Active Learning Network for Accountability and Performance
- Altus Global Alliance 283, 291–2
- AMC *see* Advanced Market Commitment
- American Bar Association 45, 88, 90, 92, 166, 325
- Amnesty International 116, 344, 347, 349
- anti-trafficking interventions/initiatives/
measures 26, 318–21, 325–6, 333–4, 342, 344, 353 *see also* human trafficking
- anti-trafficking NGOs 347, 357
- Armenia 187–8, 196
- ARWU *see* Academic Ranking of World Universities
- ASEAN *see* Association of South East Asian Nations
- Asian Development Bank (ADB) 181, 187–8, 190–2, 196–7, 202, 204–6, 460
standards 197, 205
- Asian Financial Crisis 182, 190, 192, 194–6, 200, 204, 212
- assessment process 87, 325, 329, 334
- asset classes 392, 394–5, 397–400, 414, 422, 424, 433–4
- assets 187–8, 193, 198, 275, 387, 401, 405
movable 186
- Association of South East Asian Nations (ASEAN) 322
- audit 298, 301, 303, 305, 312, 391
practices 301–3, 312
- auditees 301–2, 305
- auditors 236, 301–2, 306, 315
- Australia 104, 187–8, 366
- Australian Agency for International Development (AusAID) 283
- Australian Council for Educational Research (ACER) 60
- authoritarianism 346, 357, 359–60, 363
- authority
legitimate 282–3, 289–91, 293–4, 467
scientific 14, 16, 18, 218, 238–9, 241–2, 244
- Azerbaijan 16, 187–8
- B Corps *see* Benefit Corporations
- Bangladesh 117, 173, 431
- bankruptcy 182, 185–7, 193
corporate 23, 181, 183–4, 187, 196, 204–6
- banks 182–3, 209, 401, 409, 441
- Barnett, M. 65, 370, 372, 374
- barriers 39, 413, 425, 456–7
structural 332–3
- Barro, R. 169
- Base of the Pyramid (BoP) 397
- Belarus 16, 26, 187–8, 249, 344–64
government 346–7, 349–50, 352, 356–7, 360–2
international image 344, 347, 350, 353, 360
“Belarusian Miracle” 358
- benchmarks 70, 108, 146, 152, 167, 300–1, 368–9
- industry 393, 406, 411, 419
- beneficiaries 27, 369, 382, 385, 412–13, 421–3, 427
intended 400–1, 421
- beneficiary participation 384–5
- Benefit Corporations 405, 407, 416–17
- BERI *see* Business Environment Risk Intelligence
- Bhuta, N. 23, 35, 40, 132, 165, 169, 392
- BI *see* Business Intelligence
- Block-Lieb, S. 180–1, 193–4, 203, 205, 210–11, 216
- Board of Participating Countries (BPC) 59–61, 82
- BoP *see* Base of the Pyramid
- Bowker, G.C. 5, 135, 144, 158, 161–2
- BPC *see* Board of Participating Countries
- Brazil 58–9, 291–2, 397
- Britain *see* United Kingdom
- Bugg-Levine, A. 392, 395–6, 399, 409, 417, 425
- bureaucracy 54, 76–7, 84, 144–5, 171–5, 429, 466
- bureaucracy-driven governance 76–9
- bureaucratic delays 176
- bureaucratic efficiency 23, 166, 172–3
index 173
- Business Environment Risk Intelligence (BERI) 167–70, 174, 176–7, 179, 460
- Business Intelligence (BI) 167–8, 170–1, 173
- business models 14, 224, 396, 412–13, 416
- business schools 22, 91, 99, 104, 106
European 42, 106
rankings 98, 106
- Büthe, T. 15, 21–2, 29–31, 35, 37, 39–40, 166
- Cafaggi, F. 44
- Calhoun, M.A. 452
- Cambodia 154, 196, 331
- Camp Coordination and Camp Management (CCCM) 254
- Camp Indicator Report (CIR) 254
- camp management agencies (CMAs) 376–8, 380–1, 389
- camps 224, 366, 373, 376–8, 380–1, 386, 388–90
ghost 380
managed 372, 378, 382
unmanaged 376–8, 389–90
- capacity 64–6, 136–7, 159, 161, 192, 454–5, 457
building 412–13
- capital 35, 347, 394–400, 412–13, 415, 420–1, 427
poverty 429–32, 434
private 404, 416, 428
- capital markets 396, 398, 427
- Caple James, E. 373, 375
- Carment, D. 133–4, 136–8

- Carruthers, B.G. 4, 86, 156–7, 165, 180–1, 204–5, 211–12
- Casini, L. 28, 74, 392, 465, 472
- causal relationships 76, 147, 159, 423
- causes
- of death 222–3
 - of fragility 133–4, 137–8, 143, 154
 - of human trafficking 359–60
- CCCM *see* Camp Coordination and Camp Management
- Central and Eastern Europe 182, 185–7, 204
- Central Intelligence Agency (CIA) 141
- Centre for Educational Research and Innovation (CERI) 55, 57
- CERI *see* Centre for Educational Research and Innovation
- CESCR *see* Committee on ESCR
- CGAP *see* Consultative Group for Assistance to the Poor
- child health 225, 227
- child mortality 228, 237, 239
- children 24, 55, 217–20, 226, 238, 241, 324
- Chile 58, 291–2, 333
- China 104, 117, 191, 345, 397
- Chuang, J. 25–6, 35, 317, 327, 334, 337, 341
- CIA *see* Central Intelligence Agency
- Cingranelli, D. 152–3
- CIR *see* Camp Indicator Report
- civil liberties 36, 44, 48–9, 208, 442–3, 446, 451
- ratings 442–3
- civil society 36, 220, 284, 294, 306, 366, 386–7
- groups 37, 43, 207, 209
 - indicators as source of power for 290–3
 - organizations 36, 233, 271, 276, 310, 314, 349
- classification 5, 71–2, 91, 135–6, 154, 156, 165–6
- systems 161, 166, 186
- climate change 67, 158–9, 395, 399
- cluster approach 254, 264, 378, 383
- cluster meetings 385–6
- clustering 171–2
- clusters 120, 165, 233, 235, 254, 385–6, 389
- CMAs *see* camp management agencies
- coders 48, 152, 448
- CODHES 264–5, 269–74
- coding 152, 447–8, 450–1, 461
- coercion 229, 281, 317, 320, 328, 336, 374
- cognition 91–2, 153–4
- college rankings 42, 88, 103, 105, 108
- colleges 35, 88–9, 91, 132, 245
- Colombia 24, 117, 249–50, 255–62, 264–74, 276–9, 291–2
- government 257, 261, 264, 272
 - media 268, 270, 272
 - Ombudsman 262, 278
- Colombian Episcopal Conference 261, 277
- commensuration 87, 91–3, 95, 101–2, 150, 153–4, 215
- commerce 201, 392, 394, 415, 424, 427, 432–4
- commercial interests 35, 78–9
- commercial investments 27, 392, 403, 415, 422, 426
- Committee on ESCR (CESCR) 299–301, 304, 306, 309–12, 315
- common language 279, 285, 401, 404–5, 420
- companies 116, 124, 402–4, 406–8, 410–11, 413–15, 419–20
- portfolio 410, 423, 434
 - private 28, 438–9
- comparability 82, 202, 394, 447
- competencies 36, 56–7, 60, 70, 213, 330
- competition 39, 42, 91, 98, 101, 204, 464
- direct 419, 433
- competitiveness 181, 403, 413, 459, 466, 468
- complementary indicators 263–4
- complex phenomena 7–8, 165, 181, 444
- compliance 187–8, 263–5, 297, 305–6, 308, 323–4, 330–2
- indicators 192, 312, 314, 322–5, 338, 342
 - mechanisms 324, 326–7, 332, 335, 341–2, 468
 - scores 187–8
 - selective 346, 363
 - voluntary 468, 470, 472
- composite indicators 19, 363
- comprehensive legislation 185–6
- comprehensiveness 51, 110, 184, 447
- conditionality 195, 215, 227, 234
- Conferences of the Parties (COPs) 115, 119, 121–2, 127, 130, 323–5
- conflicts of interest 31, 242, 409, 423, 469, 471, 473–4
- Conpes documents 257–9, 278
- consensus 59–61, 134, 143, 286, 320, 334, 470
- consent 66, 253, 317, 320, 342
- tacit 78
- consistency 16–18, 23, 123, 135, 152–3, 334, 446–7
- constituents 22, 86, 97–8, 105, 107, 446, 453
- construction of indicators 91, 176, 216, 239, 283, 314
- constructivism, doctrinal 63–4
- Consultative Group for Assistance to the Poor (CGAP) 423, 431
- consumers 99, 202, 265, 417–18, 427–8
- contestation 4, 11–12, 18–20, 217–18, 239, 421, 432–4
- democratic 312–13
 - political 138, 315–16
 - public 218, 239
 - subversive 345, 361, 363
- contextualization 309, 334, 379–80
- contingencies 54, 64, 73, 75–6, 81, 193, 204
- contract enforceability 176 *see also* security, of contract rights
- contract repudiation 175–6
- contract rights, security of 175–6

- control** 11, 63, 79, 199, 278–9, 282, 327–8
Control of Corruption indicator 442
coordination 48, 67, 144, 282, 290, 366, 400
COPs *see* Conferences of the Parties
corporate activities 110–11, 114, 117, 129–30
corporate bankruptcy/insolvency 23, 181, 183–4, 187, 196, 204–6
corporate governance 12, 183, 186, 201, 209, 474
 legislation 189–90
corporate human rights responsibility 22–3, 87, 110–31, 465
 drive to measure 112–15
corporate risk 116, 118–20, 129–30
 and dangerous regions 115–21
 indicators 116
corporate social responsibility (CSR) 21, 35, 37, 112, 114, 119, 395
corporate sustainability, roadmap to 121–9
corporations 6, 92, 110–31, 168, 206, 242, 472
 multinational 41, 49, 110, 167, 171, 315, 441
corruption 3, 9, 124, 150, 170, 172–5, 441–2
 in government 175
Corruption Perceptions Index 3, 20
costs 17, 21, 31–2, 38–9, 412–13, 420, 424
Council of Europe 318, 322–3, 350–1
country assessments 331, 334, 336, 340
country narratives 331, 344, 348, 361–2
Country Policy and Institutional Assessment (CPIA) 142, 466
country rankings 53, 75
country risk data 169–70
country risk indicators 74, 120
 from academic validation to public endorsement 177–9
 ICRG *see* International Country Risk Guide (ICRG)
 lessons for indicators as governance technology 179
 origins and use 167–9
 from subjective measures to normative claims about institutional quality 169–77
coverage 20, 43, 228–30, 232–3, 235–6, 260–1, 377–8
CPIA *see* Country Policy and Institutional Assessment
credibility 18–19, 74, 108, 148, 171, 179, 454–7
credible indicators 234, 237, 293
credible threats of exit 455–7
credit rating agencies 13, 20, 28, 437, 468, 471, 473–4
creditor rights 183, 196–8, 201
creditors 186–7, 193, 196–7, 199
crime 281–3, 291, 293–4, 318, 320, 347–8, 351–3
CSR *see* corporate social responsibility
culture 121, 139, 147, 153–4, 159, 302, 356
cumulative indices 172–3
Dadush, S. 27, 38, 114, 392, 472
dangerous regions, and corporate risk 115–21
Danish Institute for Human Rights (DIHR) 116, 118
data, hard 166, 178, 329
data quality audit (DQA) 231
databases 104, 168, 173, 178–9, 184, 223
datasets 136, 148–54, 159, 168, 179, 444–5, 447
Davis, K.E. 29, 38–41, 74, 213, 220–1, 362–3, 367
death, causes of 222–3
debtors-in-management 199–200
debtors-in-possession 199
delegation 36–7, 399, 415
deliberative model of democracy 81
Deloitte 393, 400, 405, 409
demand 34–7
demand hypothesis of accountability *see* accountability, demand hypothesis of
democracy 29, 34, 49, 153, 167, 179, 443
 deliberative model of 81
democratic accountability 299
democratic contestation 312–13
Deng, F.M. 252–3, 261
Department for International Development (DFID) 133–5, 137, 141–3, 283
design of indicators, and power 281–94
despite 105
Desrosières, A. 5, 52, 66, 76, 135, 140, 144
detention 287–90, 293, 340, 345, 349, 356
 conditions 290–2
development agencies 16, 133, 140, 218, 227, 452–5, 460–1
DFID *see* Department for International Development (DFID)
diagnostic instruments 180, 183, 194, 205–6, 213, 215
diagnostic struggles 212–13
dignity 369, 379, 384–5, 389
DIHR *see* Danish Institute for Human Rights
direct competition 419, 433
disaster-affected populations/communities 369, 372–3, 380–1, 385, 391
disasters 236, 365–91
disclosure 124, 189–90, 471, 474
discretion 6, 9, 61, 153, 246, 278, 330–2
discretionary application of indicators 332–3
discrimination 118, 149, 151–2, 259, 337, 360
 gender 348, 360
 political 149–50
diseases 43, 137, 219–20, 222–5, 236, 239, 399
displaced persons/populations 24, 249–51, 255, 257–62, 267–9, 271–3, 277–8
 internally *see* internally displaced populations
displacement 251–2, 255, 257, 259, 263–4, 269–70, 272
 forced 27, 250, 265, 272, 367, 387–90

- internal 21, 250, 252–3, 256–8, 260–1, 267, 389 *see also* internally displaced populations (IDPs)
- disruption** 172–3, 214
- dissemination** 17, 61, 67, 79, 101, 156, 177
- distance** 13, 97, 114, 129–30, 286, 301–3, 373–4
- distribution**
 - in bankruptcy *see* liquidation
 - of power 4, 53
 - of resources 10, 237
- division of labor** 204–5
- doctrinal constructivism** 63–4
- “Doing Business” indicators** 470–1
- domestic aspects of indicators** 249–80, 469
- Domestic IDP indicators** 250, 264
- domestic politics** 4, 54, 209, 249, 252–3, 272, 279
- donors** 26, 226–7, 233–4, 320–1, 372, 374, 426–7
- DQA** *see* data quality audit
- DTP3 coverage** 224, 229–30, 233–7, 240, 243
 - indicator 229–30, 232–4, 245–6
- Duffield, M.** 371–3
- Dutta, N.** 27, 29, 44, 48, 50, 437, 466
- dynamism of indicators** 22, 86–109
 - data and methods of study 91
 - educational rankings 88–90
- EBRD** *see* European Bank for Reconstruction and Development
- ecology of IFIs** 203–5
- economic freedom** 150–1, 167, 442–3, 460
- economic growth** 141, 144, 158, 169–71, 177, 183, 208
- economics** 14, 32, 135–6, 145, 148, 150, 169–70
- education** 53–6, 68–70, 75–80, 103–4, 107–8, 311, 352–3
 - graduate 103–4
 - higher *see* higher education
 - legal 42, 44–6, 89, 91, 93, 98–9 *see also* law schools
 - outcomes 56, 69, 81
 - systems 54–6, 58, 70, 105–6
- educational exchange programs** 355–6
- educational policy** 53–4, 61, 67–8, 73, 77
- educational quality** 88, 90, 102, 106–7
- educational rankings** 87–8, 90, 93, 101, 103, 107 *see also* law schools, rankings
 - evolution 88–90
 - global *see* global educational rankings
- Edwards, M.** 396, 428–9
- effective enjoyment of rights** 262–5, 276, 279
- effectiveness** 137–8, 144–50, 154–5, 157–8, 160–1, 185–6, 321
 - political 148–9
 - scores 154, 158–9
- effects of indicators** 298, 300, 302, 304, 306, 308, 310
- efficacy** 215, 305, 433, 443
 - impact investment indicators 423–4
- efficiency** 16–17, 64, 171, 173–4, 176–7, 356, 369
 - bureaucratic 23, 166, 172–3
 - of government services 174–6
 - institutional 171–2
- emergencies** 194, 198, 221, 236, 258, 370, 382–4 *see also* disasters
- empirical data** 54, 56, 174, 239
- employees** 96, 121, 347, 406, 408, 410
- employers** 41, 44, 47, 89, 98, 105, 354
- employment** 55, 97, 225, 338, 353, 355
- enforcement** 41, 62, 186, 195, 197, 468
 - mechanisms 345, 353, 473
 - light 70, 74–5
- enhanced accuracy** 439, 453–7, 463
- environmental impact** 392, 395, 397, 401–3, 405, 420–1, 425–6
- environmental problems** 392, 394, 421–2, 434
- Espeland, W.N.** 5, 9, 22, 41–2, 46, 87–92, 153–4
- Ethiopia** 156
- European Bank for Reconstruction and Development (EBRD)** 181, 183–90, 192, 197, 202, 204–8, 211
 - legal indicator surveys 184–5
 - based on case studies 189–90
 - legal sector assessments 186–9
- European business schools** 42, 106
- European Humanities University** 347, 351, 357
- European Trafficking Convention** 318, 322–3, 339–40
- European Union** 5, 62, 67, 107, 322, 348, 471
- evaluation** 9–10, 15, 82, 107–8, 183, 321–3, 362–3, 445–51, 459–62
- exit** 452, 455–7
 - data 288, 293
- experimentalism** 5, 56
- expert groups** 60, 79, 81
- expertise**
 - barrier 213
 - external 80, 303
 - technical 38, 208, 213, 241, 274, 280, 306
- experts** 18–19, 81, 240–2, 306–8, 310, 322–5, 379–80
 - external 51, 80, 240
 - in-house 167–8
 - independent 60, 79, 84
 - technical 19, 242, 278, 293
- exploitation** 317–18, 336–7, 341–2, 423, 434, 456
 - trafficking-related 327, 332, 343
- extensiveness** 184–7
- external expertise** 80, 303
- external experts** 51, 80, 240
- externalities, network** 42
- faculty** 44, 91, 97, 100–1, 103, 105
 - resources 45, 89
- Failed States Index** 35, 376
- fair trade** 40, 43, 396
- Farmer, P.** 376, 380, 384

- FCO *see* Foreign and Commonwealth Office
 FDI *see* foreign direct investment
 Fearon, J.D. 134, 148, 151–2, 372
 feedback 232, 410, 424
 Feldman, M.S. 156
 FH *see* Freedom House
 finance specialists 206, 208–9
 financial institutions 23, 112, 194, 208
 financial performance 402–3, 405, 438
 financial return 38, 394, 397, 399, 409, 425
 fiscal transparency 183, 201
 Fisher, A. 3, 24, 43, 86–7, 217, 392, 437
 flows 112, 123, 168, 293, 327, 362, 399–400
 investment 427, 445–6, 457
 food 27, 225, 263, 313–14, 333, 365, 372
 for-profits 28, 394–5, 400, 403, 405, 416, 425
 forced displacement 27, 250, 265, 272, 367, 387–90
 forced labor 124, 317, 333, 335, 337–9
 Foreign and Commonwealth Office (FCO) 135, 141, 143
 foreign direct investment (FDI) 168
 foreign investors 168–9, 176–7, 208
 formal criteria 72–3, 76, 115
 Foucault, M. 65, 132, 137, 139, 315, 368, 373
 fragile states 40, 132–47, 206
 Fragile States Strategy 133–4, 136, 141, 143, 154, 156
 fragility 21, 35, 132–47, 149, 151, 153–5, 157–62
 causes/drivers 133–4, 137–8, 143, 154
 emergence of 132–40, 157
 hierarchy of 160
 indexes of 35, 132–62
 measurement 140, 142–4, 147–50, 155
 score 154–5, 159
 France 52, 102, 106, 132, 368, 373, 375
 fraud 317, 328, 336, 382
 freedom 28–30, 48–50, 349–50, 438–40, 442–3, 445–8, 450–1
 economic 150–1, 167, 442–3, 460
 Freedom House (FH) 48–51, 167, 177–8, 442–3, 446–51, 453–4, 460–2
 see also Freedom in the World ratings
 Freedom in the World ratings 28, 30, 36, 48–51, 438, 456–7, 461–2
 as test case of demand hypothesis of accountability 442–3
 as test case of supply hypothesis of accountability 460–1
 functional expert groups 60, 81
 fundamental rights 53, 69, 76, 259–60
 see also human rights
 G-7 182, 196, 204
 G-22 182, 196, 204
 GAAP *see* Generally Accepted Accounting Principles
 GAL *see* Global Administrative Law
 Gallagher, A.T. 25–6, 317–18, 321–2, 324, 326–8, 335–7, 340–2
 GAVI *see* Global Alliance for Vaccines and Immunization
 GBV *see* gender-based violence
 GC *see* Global Compact
 Gelfand, S. 403, 406
 gender-based violence (GBV) 366, 387–8
 Generally Accepted Accounting Principles (GAAP) 403, 414, 419
 generation of governance indicators 44, 48
 accountability in 437–64
 Georgia 9, 188, 196
 Germany 52, 55, 66, 68–70, 73, 102, 325
 education ministers 82–3
 Gesellschaft für Technische Zusammenarbeit (GTZ) 133, 135
 GHO *see* Global Health Observatory
 ghost camps 380
 GIIN *see* Global Impact Investing Network
 GIIRS *see* Global Impact Investment Rating System
 girls 263, 305, 340, 352–3, 360
 GIVS *see* Global Immunization Vision and Strategy
 Global Administrative Law (GAL) 4, 20–1, 63, 218, 245–6, 438–9, 468
 Global Alliance for Vaccines and Immunization (GAVI) 24, 220–1, 233–4, 240, 243
 Global Compact (GC) 21, 113–16, 119–22, 126–7, 129–31
 principles 115–16, 118–19, 121–2, 127, 129–30
 support for 121
 global educational rankings 87, 90, 93, 101, 107 *see also* educational rankings
 reactivity 103–5
 global governance *see also* *Introductory Note*
 definition 10–12
 technologies of 12–13, 20, 28, 249–50, 264, 266, 271 *see also* technologies of governance
 Global Health Observatory (GHO) 223, 239
 Global Immunization Vision and Strategy (GIVS) 233
 Global Impact Investing Network (GIIN) 38, 394, 399–402, 410, 413
 Global Impact Investment Rating System (GIIRS) 27, 393, 400, 402–4, 407–14, 419–25, 433
 global indicators 17, 23, 25–6, 28, 74, 293–4, 465–7
 accuracy 467, 469–70
 desirability of regulation 471–4
 legitimacy 469–70, 474
 regulation 74, 465–74
 regulatory implications 467–74
 scientific authority to regulatory authority 467–9
 global norms 193, 205, 211–12, 214

- Global North 370, 430, 434
 global rankings 21, 101–8
 Global Reporting Initiative (GRI) 113–15,
 121–6, 130–1, 414
 Global South 39, 180, 249, 372–3, 402, 430, 469
 power problems in design of indicators 281–94
 Goldmann, M. 5, 13, 52, 62–3, 71–4, 83–4, 256
 Goldstone, J. 137, 145–7
 good governance 24, 141, 157, 175, 177, 242, 278
 governance *see also* *Introductory Note*
 bureaucracy-driven 76–9
 model of 12
 technologies of *see* technologies of governance
 topology of 12–15, 28
 governance by information 53
 Programme for International Student
 Assessment (PISA) as 67–8
 governance indicators *see* *Introductory Note* and
detailed entries
 government policies 85, 158, 174, 282, 315,
 415, 438
 government regulation *see* regulation
 government services, efficiency of 174–6
 governmentality 5, 139, 315, 368, 373, 386
 analyses 312, 367, 373
 governmentalization of sovereignty 35, 40, 132–65
 graduate education 103–4
 GRETA *see* Group of Experts against trafficking
 in human beings
 GRI *see* Global Reporting Initiative
 Group of Experts against trafficking in human
 beings (GRETA) 322–4
 growth 20, 46, 155, 170–1, 173–4, 398, 441
 economic 141, 144, 158, 169–71, 177, 183, 208
 Grünewald, F. 380, 384–5
 GTIP (Office to Monitor and Combat
 Trafficking in Persons) 329–30, 332,
 334, 336, 341
 Guiding Principles on Internal
 Displacement 252, 258, 260
 Gurr, E.R. 145, 444–5
 Gutierrez Sanin, F. 145–6, 154–5
 Hafner-Burton, E. 345–6
 Haiti
 absence of indicators for protection 387–90
 difficulty of meeting Sphere indicators 378–80
 earthquake 365–8
 “fake” victims 380–4
 participation, not politics 384–7
 Port-au-Prince 27, 365–7, 369, 376, 378–9,
 381, 389–90
 rights-based humanitarian indicators
 post-earthquake 21, 26–7, 365–91
 spontaneous settlements and unmanaged
 camps 376–8
 Halliday, T.C. 4, 180–1, 193, 203–5, 208–12,
 216, 465
 HAP *see* Humanitarian Accountability
 Partnership
 hard data 166, 178, 329
 Hazelkorn, E. 100, 102–6, 108
 HDI *see* Human Development Index
 health 17, 40, 222–8, 234, 236–8, 243, 397–9
 public 223, 225, 227, 241, 429, 468
 statistics 222–5, 242–4
 health systems 17, 24, 217–18, 227–9, 233,
 235, 240
 performance 221, 228, 230, 238
 responsiveness 24, 227
 Heavily Indebted Poor Countries (HIPC) 24, 235
 Helleiner, E. 39
 hierarchies 8–10, 125, 154, 160, 321
 organizational 9, 153, 426, 431
 High Commissioner for Human Rights *see*
 Office of the High Commissioner for
 Human Rights (OHCHR)
 higher education 85, 87, 102, 104–7, 262,
 311–12, 359
 institutions 22, 88, 100, 103, 351
 HIPC *see* Heavily Indebted Poor Countries
 Hirschman, A.O. 455
 HIV/AIDS indicators 151, 223
 Hobbes, T. 65
 Hong Kong 58–9, 190–1
 housing 259, 263, 380, 382, 384, 395, 399
 Human Development Index (HDI) 3, 5–6, 9,
 53, 154, 409, 437–8
 human rights 3–5, 110–17, 119–21, 124–6,
 129–30, 345–6, 361–4
 abuses 110, 113–14, 116, 124, 151–3, 297,
 349–50
 corporate responsibility 22–3, 87, 110–12,
 114, 120, 122, 129–31
 indicators 35, 297–9, 302–3, 307–8,
 310–11, 315–16, 363–4
 evolving approach to 299–301
 implications of Belarus case 363–4
 renewed efforts to create international
 indicators 306
 as technologies of global governance 312–16
 law 21, 25, 35, 299, 311, 313, 345–6
 measurement 297–316, 465
 political co-optation of human rights
 protection in Belarus 344–64
 protection 9, 76, 110, 114, 345–6, 362–3
 risks 115–16, 118, 130
 treaties 25, 252, 301, 305, 307–8, 310–11, 345
 treaty bodies, troubled authority in
 international law 306–7
 violations *see* human rights, abuses
 Human Rights Risk Index 116–17
 Human Rights Watch 116, 344, 347, 349
 human trafficking 25, 344, 348, 350–2, 354–9,
 361, 364
 Belarus 344–64

- human trafficking** (*cont.*)
 case study 326–41
 causes of 359–60
 compliance indicators 322
 European Trafficking Convention 318, 322–3, 339–40
 future 341–3
 GTIP *see* GTIP (Office to Monitor and Combat Trafficking in Persons)
 indicators as indications 319–20
 indicators of performance 320–2
 TIP (trafficking in persons) 318, 348, 351–2
 indicators 326–41
 trafficking-related indicators 319–22, 326
 UN Trafficking Protocol 317, 322–7, 336–7, 339–40
 use of indicators to measure government responses 317–43
 victims 328, 332–3, 336–7, 339–40, 353, 356–7, 362
- Humanitarian Accountability Partnership (HAP)** 368–70, 372, 386
- humanitarian community** 366, 381, 385–6, 389, 391
- humanitarian indicators** 368–9
 government at a distance via indicators 371–4
 in Haiti 374–6
 rights-based 5, 21, 26–7, 365–91
 Sphere *see* Sphere
 unpacking of logics 369–71
- humanitarian intervention** 161, 368, 383
- humanitarian system** 368–70, 374, 376–8, 382, 386–9
- humanitarianism** 255, 370, 372, 374, 381, 383
- hybrid entities** 416–19
- IASC** *see* Inter-Agency Standing Committee
- ICESCR** *see* International Covenant on Economic, Social, and Cultural Rights
- ICRG** *see* International Country Risk Guide
- IDA** *see* International Development Association
- ideologies** 9–10, 49, 69–70, 129, 138, 179, 214
- IDMC** *see* Internal Displacement Monitoring Centre
- IDPs** *see* internally displaced populations
- IEA** *see* International Association for the Evaluation of Educational Achievement
- IFC** *see* International Finance Corporation
- IFIs** *see* international financial institutions
- IFRS** *see* International Financial Reporting Standards
- IHME** *see* Institute for Health Metrics and Evaluation
- Ikoyi Prison** 288–9
- ILO** *see* International Labour Office
- IMF** *see* International Monetary Fund
- immunization coverage** 220, 224, 228–31, 236, 240–2
 indicators 24, 43, 87, 217–46, 468
 in context 218–21
 demand for 224–30
 production 230–2
 as a technology of governance 237–44
 uses 232–6
- impact assessment** 321, 323, 392–3, 405, 411–13, 433
- impact investment indicators** 114, 392–434, 472
 and accountability 422–3
 altering forms of knowledge 425–8
 challenges
 to knowledge-building function 428–33
 to market-building governance function 411–14
 to regulatory governance function 420–4
 contestation of measurement 430–3
 definition of impact investing 394–7
 demands of regulating a new industry 415–18
 efficacy 423–4
 Global Impact Investing Network *see* Global Impact Investing Network (GIIN)
 Global Impact Investment Rating System *see* Global Impact Investment Rating System (GIIRS)
 and government regulation 414–18
 Impact Reporting and Investment Standards *see* Impact Reporting and Investment Standards (IRIS)
 interim indicator measures 418–20
 and knowledge production 424–5
 and market-building 394
 market weaving 398–402
 participation 420–1
 responsiveness 420–2
 and transparency 422–3
- impact investors** 395, 399–401, 408, 423–4, 434
- Impact Reporting and Investment Standards (IRIS)** 27, 38, 393–4, 402–8, 410–14, 418–25, 433–4
 development of common language for impact investing 405–7
 metrics 403, 405, 407, 411
 note on production 407
- impartiality** 16, 18, 271
- improved accuracy** 439, 453–7, 463
- in-house experts** 167–8
- incentives** 27, 38–9, 41–2, 44, 51, 305, 412–13
- independence** 22, 138, 242, 372
- independent experts** 60, 79, 84
- Index of Economic Freedom** 150–1, 460
- indexes of state fragility** 35, 132–62
- India** 106, 117, 158, 190–1, 233, 333, 397
- indicator generation** 27, 166, 240, 438–9, 454, 456–8, 462–4
- indicator generators** 27, 438–40, 452, 454–9, 462–4
- indicator producers** 13–14, 20, 27, 239, 267–8
- indicator production** 14, 22, 240, 245, 264–5, 267–70, 471

- indicators *see* *Introductory Note and detailed entries*
- Indicators of Education Systems *see* INES
- indicators of indicators 22, 111, 129
- indices 9, 160, 165, 169, 172–3, 177–9
- aggregate 168, 172, 174, 177, 179
 - cumulative 172–3
- Indonesia 58–9, 106, 120, 182, 190–1, 194, 196
- industry benchmarks 393, 406, 411, 419
- industry standards 377, 394, 402, 406
- INES (Indicators of Education Systems) 55, 79
- Networks 56–7, 59–60, 78
 - Strategic Management Group 55–6
- inferences 9, 136, 140, 151, 153, 210, 213
- infrastructure quality 176 *see also* efficiency, of government services
- INGOs (international NGOs) 26, 366, 369, 371–5, 377, 383–5, 387–8
- input legitimacy 76–81
- insolvency
- corporate 23, 181, 183–4, 187, 196, 204–6
 - law 181, 184, 187, 192, 204–5
 - systems 192–3, 195, 197–8, 201
- instability 132–3, 136–7, 141, 143, 154, 157–9, 169
- risks of 141–2
- Institute for Health Metrics and Evaluation (IHME) 243
- institution-building 133, 183, 195, 204, 206, 208–9, 214
- institutional change 169–70, 178
- institutional efficiency 171–2
- institutional quality 151, 166, 168–9, 173, 177–8
- instrumental approach 71–2
- instrumental variables 173, 178
- instruments of power 281–2
- intended beneficiaries 400–1, 421
- Inter-Agency Standing Committee (IASC) 35, 254, 381
- inter-governmental organizations/bodies 12, 15, 23, 25, 297, 318, 320
- interim indicator measures 418–20
- internal displacement 21, 250, 252–3, 256–8, 260–1, 267, 389 *see also* internally displaced populations (IDPs)
- Internal Displacement Monitoring Centre (IDMC) 255, 264–5, 267–9, 272
- internally displaced populations (IDPs) 87, 366, 377, 380, 388, 390, 469
- camps 368, 378, 380–1, 387–8
 - CODHES 264–5, 269–74
 - Colombia case study 249–80
 - governance 253–4
 - indicators 250–80
 - domestic indicators as input of global decision-making 264–6
 - entry on domestic scene 260–4
 - entry on scene 254–6
 - and global governance 250–1
 - and interaction among separate branches of power 274–9
 - and news media 266–71
 - as social mobilization 271–4
 - as technology of global governance 264
 - policy 250, 252–3, 255–6, 260–2, 271, 273, 275
 - as global governance 251–3
- International Association for the Evaluation of Educational Achievement (IEA) 56–7, 77
- international community 45, 200, 225, 339–40, 342–3, 346, 363–4
- International Country Risk Guide (ICRG) 167–8, 174–5, 179, 438, 440–3, 449–53, 460–2
- data 168–9
 - ratings 441–3, 448–9
 - as test case of demand hypothesis of accountability 440–2
 - as test case of supply hypothesis of accountability 459–60
- International Covenant on Economic, Social, and Cultural Rights (ICESCR) 25, 299, 315
- International Development Association (IDA) 134, 142, 234, 466
- International Finance Corporation (IFC) 3, 194, 197, 204
- international financial institutions (IFIs) 23, 217, 227, 237–8, 244, 246, 387
- ADB *see* Asian Development Bank (ADB)
- as diagnosticians and designers of the laws of nations 180–216
- EBRD *see* European Bank for Reconstruction and Development (EBRD)
- ecology of 203–5
- IMF *see* International Monetary Fund (IMF)
- organizational micro-politics 206–8
- professional epistemologies and competition 208–9
- World Bank *see* World Bank
- International Financial Reporting Standards (IFRS) 393, 403, 414, 419
- international image 344–7, 350, 353, 360
- international indicators 249, 271, 302, 346, 362–4
- international institutions 20, 26, 52–4, 62–4, 66, 73–5, 80 *see also* international organizations (IOs)
- International Labour Office (ILO) 320, 337, 406
- international law 4–5, 11, 63–4, 71–2, 160–1, 306–7, 334–7
- International Monetary Fund (IMF) 52, 84–5, 192–5, 201–2, 211–12, 227, 235
- international NGOs *see* INGOs
- international norms 313, 335, 339, 342, 353, 363
- international observers 350–1
- International Organization for Migration (IOM) 321–2, 347–8, 361

- international organizations (IOs)** 71–3, 180–3, 204–5, 210–13, 215–16, 329, 345
see also international institutions
- international public authority** 5, 22, 62–4, 70–2, 256
 concept 64–7
 Programme for International Student Assessment (PISA) as 68–70
- international rankings** 42, 102, 105–6
- international trade law organizations (ITLOs)** 181
- interviews** 52, 58, 60, 347, 357, 368–9, 431–2
- investees** 393–5, 400–1, 403, 405–6, 411–14, 424, 433
- investment flows** 427, 445–6, 457
- investments** 107–8, 113–14, 173, 183–4, 394–406, 415, 417–18
 commercial 27, 392, 403, 415, 422, 426
 foreign direct 168
- investors** 393–5, 397, 400–6, 409–15, 422–3, 452–3, 459–60
 confidence 394, 405, 418
- IOM** *see* International Organization for Migration
- IOs** *see* international organizations
- Iran** 117, 331, 445
- Iraq** 32, 117, 255, 268
- IRIS** *see* Impact Reporting and Investment Standards
- ITLOs** *see* international trade law organizations
- Jacobsson, B.** 77, 310–12
- Jamaica** 25, 283–5, 288–90, 292–3
- Japan** 104, 106, 123, 190–1, 320, 333
- journalists** 49, 51, 76, 84, 329, 444, 461–2
- judicial activism** 274, 277, 279
- judicial reforms** 194–5
- justice** 3–4, 39, 64, 71, 76, 195, 375
 active indicators of 284–90
 power problems in design of indicators 281–94
- Kaufmann, D.** 14, 17, 176–7, 441–2
- Keefer, P.** 169–71, 174–8
- Kingsbury, B.** 3–4, 14–15, 74, 367, 437–9, 468–9, 472
- Knack, S.** 169–71, 174–8, 441
- knowledge** 8, 17–18, 57, 59, 75–6, 85–6, 136–7
 altering forms of 425–8
 production 327, 332, 342–3, 393–4, 424
 and impact investment indicators 424–5
- knowledge-building function, challenges to** 428–33
- knowledge-effects** 158
- Korea** 190–1, 194
- La Strada** 347–8, 351, 354–7
- labor** 114, 116, 124, 172–3, 317, 336–7
 division of 204–5
 forced 124, 317, 333, 335, 337–9
- Lagos** 287–8, 292
- Lampland, M.** 5, 15, 156
- language** 27, 69, 255, 302–3, 310–11, 319, 385–6
 common 279, 285, 401, 404–5, 420
 of indicators 326, 341
 of mathematics 303
 of quantification 301, 303
- Latour, B.** 5, 11, 14, 19, 86, 123, 157
- Latvia** 58–9, 188
- law and order index** 174
- law reform** 206, 212–13, 215
- law schools** 3, 42, 45, 47–8, 87–93, 96–7, 132
 administrators 91, 97–8
 deans 44–5, 96
 rankings 22, 35, 44–8, 87–91, 95, 98, 100–2
- laws of nations** 86, 180–216, 465
- league tables** 8, 100, 102–3, 105–8, 228
- legal basis** 59, 66, 72, 80, 351
- legal change** 181, 189, 211–13, 215
- legal departments** 180–1, 184, 187–8, 194–6, 206–7, 209–11, 214–15
- legal education** 42, 44–6, 89, 91, 93, 98–9
see also law schools
- legal frameworks** 54, 62, 76, 79–80, 193, 336–8, 473–4
- legal indicator surveys** 184–5
 based on case studies 189–90
- legal indicators** 184–5, 189, 207, 214
- legal regimes** 73, 75, 79, 83–5, 195, 317, 467
 NPAs 79, 81, 83–4
- legal sector assessments** 186–9
- legal systems** 16, 180–1, 186, 193, 205, 214–15, 466
- legal transition indicators** 186–7
- legal yardsticks** 86, 180–216, 465
 EBRD *see* European Bank for Reconstruction and Development (EBRD)
- IMF** *see* International Monetary Fund (IMF)
- politics and theory** 202–3
- Reports on the Observance of Standards and Codes (ROSCs)** 183, 193, 200–3, 207–9, 211
- varieties of** 182–3
- World Bank** *see* World Bank
- legality** 62, 74, 145, 200, 214, 336, 439
- legislation** 202, 258–9, 277, 352, 471, 473
 comprehensive 185–6
- legitimacy** 62–4, 75, 143–50, 157–61, 209–11, 274–7, 466–7
 global indicators 469–70, 474
 input 76–81
 national policy assessments (NPAs) 75–84
 output 81–2
 personal 79
 political 138, 147, 149
- legitimate authority** 282–3, 289–91, 293–4, 467
- legitimation** 67, 210–11, 345–6
- Lesser, W.** 304
- leverage** 27, 158, 183, 196, 207, 209, 416
- liberal democracy** 48–9, 62
- liberty** 48, 62–4, 66–7, 76, 118, 284

- light enforcement mechanisms** 70, 74–5
liquidation 185–7, 193, 199
local networks 126–9, 131
location 94, 97, 108, 116, 120, 125–6, 377–8
logframes 374
Lukashenko, Alyaksandr 348–51, 353, 357–60, 362–4

magazines 20, 35, 46, 48, 88–90, 101, 443–4
Malaysia 104, 182, 190–1, 333
managed camps 372, 378, 382
management 57, 112, 159, 198–9, 224, 290, 391
Maplecroft 116
MAR *see* Minorities at Risk
March, J.G. 9, 98, 156, 165
market-based solutions/systems 399, 415, 418, 421
market-building 394, 411, 414, 434
 and impact investment indicators 394
market logic 425–6, 428, 433
market weaving 398–402
marketing 96–7, 245, 311
marketization 426–30
markets 34–5, 37–8, 40–1, 208–9, 397–400, 403, 409–11
Martens, K. 55–7, 67–8, 77–8
Martin, M. 427
Maternowska, C. 384
mathematics 56, 59–60, 103, 303, 312
Mattli, W. 15, 30–1, 35–7, 39–40, 166, 419
Mauro, P. 167–8, 170–4, 177
MCC *see* Millennium Challenge Corporation
measurable indicators 254, 310, 370
measurement 29–32, 42–3, 122–3, 129–30, 142–4, 148, 332–4
 contestation of 430–3
 politics of 333–5
 standards 29–32, 34, 36, 38, 40–1
measures, tendency to become targets 304–5
media 75, 84, 88, 95, 266–7, 270–1, 443
 Colombia 268, 270, 272
 news 20, 250, 264, 266, 271, 280
mediation of the reception of indicators 69, 91, 97–8, 100, 105–6
Merry, S.E. 3–5, 29, 74, 86, 93, 213–15, 367
meta-indicators 328, 332
Mexico 158, 196, 291–2
micro-politics, organizational 206–8
Migdal, J. 138–9
Millennium Challenge Corporation (MCC) 3, 220, 234–6, 442–3, 452–3, 456–7, 459–61
 Indicators 442–3, 453
minimum standards 326, 328–9, 331–3, 362, 367–8, 377, 379–80
Minorities at Risk (MAR) 28, 149–51, 439–40, 447–9, 451, 461
 documentation 444, 447–51
 indicators 445, 461, 463
 as test case of demand hypothesis of accountability 444–5
 as test case of supply hypothesis of accountability 461
minority shareholders 185–6
mobilization, social 149, 250, 264, 271, 274, 276, 279
models 12, 20, 22, 50, 139, 346, 473
 of supply and demand 31–2, 50
modern slavery *see* human trafficking
money 11, 16, 42, 88, 97, 166, 168
morbidity 223, 225, 233, 239
Most Significant Change (MSC) model 431
MSC *see* Most Significant Change (MSC) model
multi-stakeholder conceptual model 33
multinational corporations 41, 49, 110, 167, 171, 315, 441
multiple indicators 98–9, 108, 470
Munck, G. 144–5, 147, 155
Murray, C.J.L. 230, 242–4

naming 8, 35, 334
narratives, country 331, 344, 348, 361–2
national human rights institutions (NHRIs) 308, 310, 314
national level 63, 66, 80, 104, 200, 224, 230–1
national policy assessments (NPAs) 5, 22, 52, 71–3, 75, 79–81, 83–5
 implementation 79–80
 legal regime 75–84
 consequences of input legitimacy 79–81
 consequences of output legitimacy 83–4
 as possible avant-garde of an information-based multilevel democracy 85
 and standard instrument definition 72–5
national rankings 100, 106–7
nationalization potential 176 *see also* security, of property rights
Natsios, A. 429–31, 466
Nepal 226, 238
network externalities 42
networks 5, 14, 21–3, 56–7, 111–12, 120–5, 129–31
 of constituents 22, 86
 failure 123–4
 spatialities 121–9
news media 20, 250, 264, 280
 and internally displaced populations indicators 266–71
newspapers 90, 101, 266–9, 443–4
NGOs (non-governmental organizations) 26, 266, 271–4, 329–30, 356–7, 366–7, 373–5
NHRI *see* national human rights institutions
Nigeria 117, 283, 287–90, 292–3, 333
Nigeria, Lagos 287–8, 292
non-compliance 121, 236, 345–6, 353, 420
non-governmental organizations *see* NGOs
non-state actors 24, 39, 45, 160, 217, 249, 472

- norm-vessels** 394, 424, 432–3
normalization 139, 375, 384
norms 12, 30–1, 138–9, 181–3, 204–5, 210–11, 426–7
 global 193, 205, 211–12, 214
 international 313, 335, 339, 342, 353, 363
North, D. 145, 169–70
North Korea 117, 331
Norwegian Refugee Council (NRC) 264–5, 267–8
NPAs *see* national policy assessments
NRC *see* Norwegian Refugee Council

objective indicators 170, 178, 304
objectivity 5, 14, 16, 74–5, 143–4, 301–4, 332–4
OCHA *see* Office for the Coordination of Humanitarian Affairs
OECD (Organization for Economic Cooperation and Development) 13, 22, 52–61, 70, 73, 75–8, 84–5
 Council 55, 58–60, 80
 PISA *see* Programme for International Student Assessment
 Secretariat 55–61, 77, 79, 82–3
Office for the Coordination of Humanitarian Affairs (OCHA) 365, 369, 377
Office of the High Commissioner for Human Rights (OHCHR) 3, 25, 265, 269, 301–2, 306, 312–15
 indicators initiative 307–11
Office to Monitor and Combat Trafficking in Persons *see* GTIP
OHCHR *see* Office of the High Commissioner for Human Rights
Olson, M. 170, 456
online surveys 365, 367–9, 377, 379–80, 383, 385, 387
operational agencies 286, 289–90
ordinal ranking 8, 153–4, 157, 201–2, 205
Organization for Economic Cooperation and Development *see* OECD
organizational hierarchies 9, 153, 426, 431
organizational micro-politics 206–8
organizational politics 206, 211
outcome indicators 149–50, 310
outcomes 14, 20, 35, 148–9, 171, 313–14, 374
 of education 56, 69, 81
 political 134, 138, 145, 255
output legitimacy 81–4
outputs 17, 48, 70, 321, 383, 423–4, 429–30
outsiders 356, 440, 445, 448, 451–2, 457–8, 463

Pakistan 117, 190–1, 405
Pan-American Health Organization 231, 276
panoramas 155, 157–8
Papua New Guinea 283, 286, 290, 293
participation 369–71, 384–7, 439–40, 448–9, 458–9, 461–4, 472
 impact investment indicators 420–1

Patrick, S. 133, 135–6, 140–1
perceptions 14, 53, 134, 146–9, 151, 185–6, 241
performance 9, 26, 53–4, 81–2, 93–4, 281–3, 290–1
 financial 402–3, 405, 438
 indicators 56, 123–4, 227, 234–7, 246, 320–1, 402
personal legitimacy 79
Peru 58, 196, 291–2
PGB *see* Programme for International Student Assessment (PISA), Governing Board
philanthropy 27, 392, 394–5, 399, 415, 424–7, 432–4
Philippines 117, 190–1
PISA *see* Programme for International Student Assessment
Pistor, K. 23, 29, 41, 48, 86, 148–9, 214
placement 89, 93, 245
pledges 130, 185–6
police 147, 281–3, 286–7, 289–91, 293–4, 357, 416
Police Station Visitors Week 290–3
political contestation 138, 315–16
political discrimination 149–50
political-economic conceptual model 29–51, 466
 applications 45–50
political effectiveness 148–9
political institutions 37, 43–4, 48, 137–8, 147, 175
political legitimacy 138, 147, 149
political order 35, 143–5, 149, 151–3, 157, 159, 161
 calculability 132–62
political outcomes 134, 138, 145, 255
political rights 36, 44, 48–9, 85, 116, 442–3, 446
Political Risk Services (PRS) ratings 440, 454
political risks 3, 41, 49, 63, 171, 449, 460
 ratings 446, 448–9
political stability 133, 169–70, 172–3, 177, 209, 441
 index 173
Political Terror Scale 150–3
politics 10, 25–6, 30–1, 169–70, 202, 206, 215–16
 domestic 4, 54, 209, 249, 252–3, 272, 279
 of measurement 333–5
Polity IV 150–1
Port-au-Prince 27, 365–7, 369, 376, 378–9, 381, 389–90
Porter, T.M. 5, 16, 18, 39, 144, 301, 303
portfolio companies 410, 423, 434
potable water 365–6, 375, 390
poverty 133–4, 141, 194–5, 235, 360, 395–6, 429–30
 capital 430
Poverty Reduction Strategy Papers (PRSPs) 235
power 4, 53–4, 63–5, 208–10, 236–40, 281–3, 315–17
 balance of 19, 242

- and design of indicators 281–94
- indicators as source of for civil society 290–3
- instruments of 281–2
- relations 10, 13, 50, 204, 386
- relationships of 32–3, 46, 50, 65
- predictability** 143, 197, 254–5, 275, 394, 429
- predictions** 154, 445–6, 451–2, 461–2
- Prest, S.** 133–4, 136
- pretrial detention** 287–9, 292–4
 - indicator 289, 293
- prioritization** 236, 313–14, 375, 397, 421–2
- PRIs** *see* program-related investments
- prisons** 65, 287–9, 293, 345
- private actors** 11, 20, 407
- private bodies/entities** 31, 36–7, 40, 180, 421, 467
- private capital** 404, 416, 428
- private companies** 28, 438–9
- private indicators** 36–7, 41, 48, 86, 149, 467, 469
 - of country risk
 - from academic validation to public endorsement 177–9
 - lessons for indicators as governance technology 179
 - origins and use 167–9
 - from subjective measures to normative claims about institutional quality 169–77
 - reconstruction for public purposes 165–79
- private investors** 209, 459–60
- private regulation** 30, 37–8
- private sector** 36–7, 39, 47, 207, 415, 445
- process indicators** 25, 262
- proficiency levels** 57, 59, 61
- program-related investments (PRIs)** 417
- Programme for International Student Assessment (PISA)** 13, 40, 90, 466, 468
 - as challenge for public law 62–7
 - development of legal framework for 71–84
 - as example of bureaucracy-driven governance 76–9
 - genesis 54–8
 - as governance by information 67–8
 - Governing Board 58–61, 74, 76, 79–80, 82, 84
 - impact 67–71, 73
 - implementation 54, 79, 81
 - as international public authority 68–70
 - legal reconstruction 52–85
 - legitimacy 54, 76, 79
 - in operation 58–62
- project management** 374, 377, 385
- promulgators** 3, 10, 13–14, 129, 245, 467–71
- property** 186, 198, 252
 - rights 35, 147, 150, 195, 441
 - security of 23, 174–6
- prosecutions** 283, 325, 328–31, 334, 353, 356, 362
- prosecutors** 281–3, 287–9
- prospective students** 44, 46–7, 89, 91, 97–8
- prostitution** 26, 280, 317, 337–8
- proxies** 142, 144, 147–8, 154–5, 171, 173–4, 217–18
- PRS Group** 28, 168, 438, 440–1, 448–50, 452, 459–62
- PRSPs** *see* Poverty Reduction Strategy Papers
- public authority** 20, 52–4, 62–7, 71–3, 76, 83–5, 472
 - acts of 71, 73, 75, 81
 - international 5, 22, 62–4, 66–8, 70–2, 256
- public contestation** 218, 239
- public endorsement** 177–9
- public health** 223, 225, 227, 241, 429, 468
- public law** 22, 53, 62, 70, 72, 80
 - function of 62–4
 - indicators as challenge for 52–4 *see also* Programme for International Student Assessment (PISA)
- public regulation of global indicators** *see* regulation, of global indicators
- QSWUR (QS World University Rankings)** 87, 100–2, 104–7
- qualitative assessments** 142, 304, 444, 448, 451
- qualitative rankings** 187–8
- quality, institutional** 151, 166, 168–9, 173, 177–8
- quantification** 4–5, 27–8, 140–4, 151, 153, 158–9, 303
 - language of 301, 303
- quantitative indicators** 98, 286, 299, 444
- questionnaires** 60–1, 123, 168, 323, 329–30
- rank ordering** 154–6, 160–1
- rankings** 4–9, 44–6, 87–108, 154, 156–7, 186–8, 202 *see also* league tables
 - business schools 98, 106
 - college 42, 88, 103, 105, 108
 - educational *see* educational rankings
 - global 21, 101–8
 - global educational *see* global educational rankings
 - international 42, 102, 105–6
 - law schools 22, 35, 44–8, 87–91, 95, 98, 100–2
 - national 100, 106–7
 - ordinal 8, 153–4, 157, 201–2, 205
 - QSWUR (QS World University Rankings) 87, 100–2, 104–7
 - qualitative 187–8
 - tier 330–1, 333–4, 343
- rape** 328, 366, 388
- rating agencies** 13, 20, 28, 207, 403, 471, 473–4
- ratings** 49–50, 152, 408–11, 419–22, 440–1, 446–52, 460–1
 - civil liberties 442–3
 - Freedom in the World *see* Freedom in the World ratings
 - political risks 446, 448–9
 - preliminary 446–7

- rational choice theory 65–6, 170
- rationality 91, 213, 255, 275, 278, 432
- raw data 6–7, 18, 23, 179, 220, 272, 422
- raw scores 447, 450
- RBM *see* results-based Management
- reactivity 5, 46, 48, 88–9, 91–2, 95–7, 103
- reason-giving 15, 20, 445–6, 450–1, 454, 458–9, 461–2
- reception of indicators, mediation of 91, 97–100, 105–6
- reconstruction of private indicators for public purposes 165–79
- recursivity 203, 211–14
- red tape 171–3
- reforms 16, 68–70, 106, 194–5, 208, 213–15, 345–6
- refugees 150, 250–1, 254, 257, 267, 270
- regions 22–3, 111–12, 118–20, 125–6, 129–31, 192, 230–2
- regulation 10–12, 20–1, 26–8, 30–1, 36, 40, 414–20
 - of global indicators 74, 465–74
 - desirability 471–4
 - private 30, 37–8
 - threat of 39, 43
- regulators 11, 197, 276, 418, 420, 467, 469–70
- regulatory implications of global indicators 467–74
- relationships of power 32–3, 46, 50, 65
- reliability 18, 21, 34, 59, 151–2, 330, 332–3
- reliance 16–18, 130, 150–1, 180, 213, 267, 303
- remand prisons 287–8, 293
- repatriation 250–1, 357
- repeatabilities 470
- Reports on the Observance of Standards and Codes (ROSCs) 183, 193, 200–3, 207–9, 211
- repressive regimes 345–6, 363
- reputation 45, 75, 89, 93, 98–100, 454–5, 460–2
- resources 10–11, 104, 145–6, 176–7, 204, 418, 457–8
- respondents 116, 185, 369, 377, 379–81, 383
- responsiveness 48, 168, 420–2
- results-based Management (RBM) 254, 374
- return, financial 38, 394, 397, 399, 409, 425
- review 245–6, 323–5, 445–8, 450–1, 453–4, 458–9, 461–4
- rights 25, 186, 238–9, 258–63, 299–300, 312–14, 387–9
 - effective enjoyment of 262–3, 265, 279
 - fulfillment 25, 297, 299, 302, 305, 313
 - fundamental 53, 69, 76, 259–60
 - human *see* human rights
- rights-based humanitarian indicators 5, 21, 26–7, 365–91
- risk assessment 141, 207, 430, 437, 444, 446, 451
- risk profiles 118, 167
- risk ratings 118, 441, 446, 448–50, 452
- risk-to-people 119, 122, 129–31
- risks 82–4, 113–14, 116–20, 129–33, 156–9, 175–9, 244–6
 - of instability 141–2
 - political 3, 41, 49, 63, 171, 449, 460
- Rockefeller Foundation 3, 14, 38, 395–402, 405, 409, 415
- Rodriguez, C. and D. 258–60, 263
- ROSCs *see* Reports on the Observance of Standards and Codes
- Rosga, A. 5, 25, 35, 68, 255, 297, 367
- rotavirus vaccine 219, 233, 237
- routine immunization *see* immunization coverage
- rule-demanders 32–4, 46, 49
- rule-makers 32–4, 38, 46, 49
- rule of law 7, 9, 23, 41, 174–5, 195, 283
 - index 7, 166, 174, 176
- Russia 117, 196, 233, 349, 358
- Russian Federation 58–9, 188
- SACs *see* Standards Advisory Councils
- safety 39–40, 249, 281–94, 469
 - active indicators of 284–90
- safety and justice indicators, power problems in
 - design of 281–94
- Salmi, J. 100, 102, 105–6, 108
- sanctions 149, 331, 350, 362, 471, 474
 - mechanisms 327, 330–1
- Saroyan, A. 100, 102, 105–6, 108
- Satterthwaite, M.L. 5, 25–6, 35, 68, 255, 297, 299
- Sauder, M. 22, 35, 38, 41–2, 45–6, 48, 88–92
- scatter-plots 284–5, 287
- Schlichte, K. 138–9
- Schuller, M. 366, 373, 375, 378, 381, 386, 388–9
- scientific authority 14, 16, 18, 218, 238–9, 241–2, 244
- scientific community 13–14, 242, 245
- scientific expertise 21, 60, 239, 242, 393, 473
- scores 44–5, 48–9, 90, 142, 155, 158–9, 290–2
 - raw 447, 450
 - star-based 393, 402, 408, 424
- scrutiny 20, 74, 94, 166, 181, 218, 227
- sector indicators 263–4, 367
- security 27, 116, 132–3, 137, 139–41, 146, 245
 - of contract rights 175–6
 - interests, timely registration 185–6
 - of property rights 23, 174–6
 - risks 136, 140–1, 159
- selection 17, 81, 156, 166, 171, 308–9, 411
- selective compliance 346, 363
- selectivity 45, 89, 93, 363
- self-assessment 116, 118–19, 198, 201
- self-regulation 12, 21, 26, 33, 50
- semiotics 87, 110–31, 465
 - case study *see* corporate human rights responsibility
- settlements, spontaneous 376–8, 385
- shareholders 186, 189, 417–18, 421–2
 - minority 185–6

- shelter 27, 79, 333, 340, 366, 371–2, 383
 - space indicators 379
- Sierra Leone 154, 196, 283, 293
- similarity 111–12, 120, 123, 130, 157, 474
- simplification 6–8, 87, 92, 203, 346, 362, 438
- Singapore 106, 165, 190–1, 228, 320
- SIR *see* Standards and Indicators Report
- skills 57, 59, 123, 134, 213, 229, 282
- slavery 317, 336–7, 339, 354, 356–7, 375, 384 *see also* human trafficking
- smallpox 225–6
- Snidal, D. 11, 36
- social balance sheet 395–6, 412–13, 421
- social entrepreneurs 396
- social goods 415, 418, 420
- social groups 45, 146, 308, 319
- social mobilization 149, 250, 264, 271, 274, 276, 279
- social orders 136–7, 145, 157
- Social Returns on Investment (SROI) 426
- social value 406, 427–8, 432
- Somalia 117, 391
- sources 108, 150–1, 166, 186–9, 266–70, 329–30, 447–9
- South Africa 117–18, 260, 274
- sovereignty 160–2, 253, 255, 325, 342, 372
 - governmentalization 35, 40, 132–65
- spatialities, networks 121–9
- Sphere 26, 367–71, 377–80, 383–5, 387–9
 - development of project 369–71
 - Handbook* 27, 371, 379, 390
 - indicators 26–7, 367, 378–9
- Sphere, Minimum Standards in Disaster Response 367–8, 374, 383
- spider graphs 188–9
- spontaneous settlements 376–8, 385
- SROI *see* Social Returns on Investment
- stability, political 133, 169–70, 172–3, 177, 209, 441
- stakeholders 30–3, 38–9, 43, 46–7, 50–1, 105, 421–3
 - feedback 424, 426
- standard instruments 22, 54, 71–3, 85
 - definition 72–5
 - as key category 71–2
- standard-setting 4, 15–16, 28, 30–1, 40, 112, 184
- standardization 30, 40, 166, 298, 301, 310, 362
- standards 15, 30–1, 39–44, 310–12, 367–71, 379–80, 418–20
 - minimum 326, 328–9, 331–3, 362, 367–8, 377, 379–80
- Standards Advisory Councils (SACs) 409–10
- Standards and Indicators Report (SIR) 254
- standing 15, 97, 99, 106, 115–16, 256, 290
- star-based scores 393, 402, 408, 424
- Starr, L. 144, 158, 161–2
- stars 402–3, 408, 430–1
- state failure 21, 141, 146, 149, 155, 157, 375
- state fragility *see* fragility
- stateness 136, 139–40, 154, 157
- Stevens, M. 9, 11, 92–3, 151, 153, 158, 346
- strategy 57–8, 60–1, 63, 78, 80–1, 96–7, 105–6
 - fragile state 133–4, 136, 141, 143, 154, 156
- Strategy Framework for the Assessment and Treatment of Fragile States 137, 145–7
- students 56–7, 61, 68–70, 82, 91–7, 105–7, 359–60
- subjective indicators 171, 173
- subsidies 273, 415–16, 419, 425
- subversive contestation 345, 361, 363
- success 9, 90, 178–9, 235, 276, 320–1, 432
- suppliers 18, 21, 38, 47, 50–1, 413, 427–8
- supply 37–40
- supply chains 122, 333, 414
- supply hypothesis of accountability *see* accountability, supply hypothesis of
- supply of indicators 17, 31, 33, 38–9
- survey data 74, 78, 231–2, 243, 294
- surveys 20, 45, 47–8, 56–7, 73–6, 83–4, 184–6
 - online 365, 367–9, 377, 379–80, 383, 385, 387
- Taiwan 106, 182, 190–1
- target countries 331, 455, 457
- target groups 49, 228, 237
- target populations 224, 231–2, 241, 406
- targets 438–40, 452, 454–7
 - and measures 304–5
 - ultimate 32, 43–4
- taxonomies 28, 406, 419, 423, 467, 469
- technical expertise 38, 208, 213, 241, 274, 280, 306
- technical experts 19, 242, 278, 293
- technocrats 211, 276–7, 279
- technologies of governance 28–9, 179–80, 249–50, 264–6, 279–81, 465–9, 471–2
 - human rights indicators as 312–16
 - immunization coverage indicators as 237–44
 - indicators as 10–21
- terrorism 172–3, 183, 213
- Thailand 58–9, 182, 190–1, 226
- threats of exit, credible 455–7
- tier rankings 330–1, 333–4, 343
- timely registration of security interests 185–6
- TIMSS *see* Trends in International Mathematics and Science Study
- topology of governance 12–15, 28
- trafficking
 - in human beings *see* human trafficking
 - in persons *see* human trafficking
- Trafficking Protocol 317, 322–7, 336–7, 339–40
- trafficking-related exploitation 327, 332, 343
- trafficking-related indicators 319–22, 326
- Trafficking Victims Protection Act of 2000 (TVPA) 318, 327–31, 334, 336, 339–40
- Trafficking Victims Protection Reauthorization Act of 2003 (TVPRA) 318, 329–30, 332, 335, 338
- transaction costs 34, 404, 406, 412

- transparency** 15–16, 290–1, 404–5, 439–40, 448–51, 458–9, 461–2
 fiscal 183, 201
 and impact investment indicators 422–3
Transparency International 3, 14, 20, 460
transposition 23, 166, 169, 171, 174, 179, 469
trends 52, 69, 92, 160–1, 170, 221, 324
Trends in International Mathematics and Science Study (TIMSS) 56, 82
tuberculosis 226, 236, 467
TVPA *see* Trafficking Victims Protection Act of 2000
TVPRA *see* Trafficking Victims Protection Reauthorization Act of 2003
- Ukraine** 188–90, 349, 358
ultimate targets 32, 43–5
uncertainties 9, 19, 112, 151, 153, 165, 239
uncertainty absorption 9, 153, 165, 346, 426
UNCITRAL (United Nations Commission on International Trade Law) 181, 184, 193–4, 197, 204–5, 209, 211
unconstitutional states of affairs 260, 262–4, 275
UNDP (United Nations Development Program) 3, 53, 133, 220, 227, 437–8
UNHCR (UN High Commissioner for Refugees) 150, 251, 253–5, 257, 262, 264–70, 272
UNICEF (United Nations Children's Fund) 217, 220–1, 226, 230–2, 234, 240, 369
unilateralism 26
United Kingdom 63, 68, 102, 159, 233, 274, 418–19
 Foreign and Commonwealth Office (FCO) 135, 141, 143
United Nations (UN) 24–5, 205, 251–4, 265, 298, 347–9, 437
 Children's Fund *see* UNICEF
 Commission on International Trade Law *see* UNCITRAL
 Development Program *see* UNDP
 Global Compact *see* Global Compact (GC)
 High Commissioner for Human Rights *see* Office of the High Commissioner for Human Rights (OHCHR)
 High Commissioner for Refugees *see* UNHCR
 indicators in critical perspective 297–316, 465
 Office on Drugs and Crime *see* UNODC
 Relief and Reconstruction Agency *see* UNRRA
 Trafficking Protocol 317, 322–7, 336–7, 339–40
United States 54–5, 133, 140–1, 219, 331–2, 334–5, 338–9
 government 49, 326, 329, 333, 337–8, 462
 GTIP (Office to Monitor and Combat Trafficking in Persons) 329–30, 332, 334, 336, 341
 State Department 26, 143, 331–2, 334–6, 343–4, 347–8, 361–2
- TIP (trafficking in persons)** 318, 348, 351–2
 indicators 326–41
Trafficking Victims Protection Act of 2000 (TVPA) 318, 327–31, 334, 336, 339–40
Trafficking Victims Protection Reauthorization Act of 2003 (TVPRA) 318, 329–30, 332, 335, 338
USAID *see* USAID
universities 41, 88, 91–2, 101, 103–8, 203, 359
unmanaged camps 376–8, 389–90
unobservable variables 147–8
UNODC (UN Office on Drugs and Crime) 318, 320, 325–6, 339, 344, 347, 352
UNRRA (United Nations Relief and Reconstruction Agency) 251
USAID (US Agency for International Development) 132–7, 141–51, 154, 156–8, 234, 400–1, 409–10
Index 147–9, 151–61
 absorption of uncertainty, commensuration, aggregation 150–9
 origins 140–4
 theorizing the state, prototyping strength, and counting what can be counted 144–50
use of indicators 40–3
 country risk 167–9
 to measure government responses to human trafficking 317–43
users 21, 31–4, 46–51, 166–7, 244–5, 439, 463–4
 hedging by 452–3
 potential 32, 40, 46, 309, 459
 unwillingness to undermine indicators 453–4
USN (US News and World Report), law school rankings 22, 30, 35, 38, 45–8, 87–91, 94–101
- vaccinations** 218, 220, 224, 229–30, 232, 237
 see also immunization coverage
vaccine-preventable diseases 219–20, 224, 230, 232–3, 239–40
vaccines 218–19, 221, 224, 229, 232–4, 237–40
 new 24, 219, 221, 230, 233–4
validity 21, 99, 112, 151, 171, 177, 213–14
variables
 instrumental 173, 178
 unobservable 147–8
variance 123, 291
verification 12, 152, 261, 272, 298, 330
Verkuilen, J. 144–5, 147, 155
victims of trafficking 328, 332–3, 336–7, 339–40, 353, 356–7, 362
violence 11, 35, 136, 139, 174, 256–8, 280
visibility 88, 103–4, 108, 258, 461
voluntary compliance 468, 470, 472
von Bogdandy, A. 40, 52, 62–3, 71–4, 83–4, 256, 468
vulnerability 133, 158, 200, 317, 373, 381–2

- Wall Street Journal** 99, 228
Washington Consensus 23, 170
watch list 14, 141, 330
water 27, 77, 91, 380, 383–4, 428, 432–3
 potable 365–6, 375, 390
Weber, M. 62, 64, 138–9, 149, 183, 274, 277
websites 88, 90, 101, 116, 133, 179, 444–5
weightings 6, 19, 102, 363, 409–11, 421
WGIs *see* Worldwide Governance Indicators
WHO *see* World Health Organization
WJP *see* World Justice Project
WME ethical index 112–13
Wolf, K.D. 55–7, 77–8
women 278, 280, 311, 321, 352, 354, 360
World Bank 15–16, 141–2, 170, 176–7, 192–207, 226–8, 234–7
 Assessment Template 198–200
 Insolvency Initiative 196, 198
 principles and guidelines for national insolvency regimes 196–8
World Health Organization (WHO) 24, 53, 150, 217–34, 236–7, 239–44, 472
 health statistics expert 221–4
 World Health Report 24, 53, 223, 227–8
 World Health Statistics 223, 239
World Justice Project (WJP) 166
Worldwide Governance Indicators (WGIs) 441–3, 456, 459
writs 260–3, 265, 272, 276
YWCA 347–8, 351