

December 2010

Accountability in the Generation of Governance Indicators

Nikhil K. Dutta

Follow this and additional works at: <https://scholarship.law.ufl.edu/fjil>

Recommended Citation

Dutta, Nikhil K. (2010) "Accountability in the Generation of Governance Indicators," *Florida Journal of International Law*: Vol. 22: Iss. 3, Article 3.

Available at: <https://scholarship.law.ufl.edu/fjil/vol22/iss3/3>

This Article is brought to you for free and open access by UF Law Scholarship Repository. It has been accepted for inclusion in Florida Journal of International Law by an authorized editor of UF Law Scholarship Repository. For more information, please contact kaleita@law.ufl.edu.

ACCOUNTABILITY IN THE GENERATION OF GOVERNANCE INDICATORS

*Nikhil K. Dutta**

I.	INTRODUCTION	403
II.	AN OVERVIEW OF GOVERNANCE INDICATORS	405
	A. <i>What Are Governance Indicators?</i>	405
	B. <i>Who Develops Governance Indicators?</i>	408
	C. <i>Uses of Governance Indicators</i>	409
	D. <i>Motivations for the Use of Governance Indicators</i>	413
	1. Proximate Motivations.....	413
	2. Fundamental Motivations	414
	3. Simplification and Uniformity	414
	4. Bolstering Credibility and Concealing Judgment	415
III.	CONCEPTS AND CONCERNS	416
	A. <i>Indicator Users, Generators, and Targets</i>	417
	B. <i>Accuracy</i>	418
	C. <i>Accountability</i>	419
	D. <i>Global Administrative Law</i>	420
IV.	THE DEMAND HYPOTHESIS OF ACCOUNTABILITY	422
	A. <i>The Contribution of Accountability</i>	422
	B. <i>The Provision of Accountability by Generators</i>	424
	C. <i>Method of Evaluating the Hypothesis</i>	425
V.	THREE TEST CASES.....	425
	A. <i>International Country Risk Guides, PRS Group</i>	426
	1. Institutional Background.....	426
	2. Users and Targets.....	427
	B. <i>Freedom in the World Ratings, Freedom House</i>	428
	1. Institutional Background.....	428
	2. Users and Targets.....	429
	C. <i>Minorities at Risk Dataset, MAR Project</i>	430

* Law Clerk, Hon. Stewart R. Dalzell, Eastern District of Pennsylvania. J.D. 2010, New York University School of Law; M.P.A. 2007, Woodrow Wilson School of Public and International Affairs, Princeton University; A.B. 2002, Harvard College. I would like to gratefully acknowledge the advice and encouragement of Christopher Bradley, Kevin Davis, Angelina Fisher, Benedict Kingsbury, and Richard Stewart. All errors that remain are my own.

1. Institutional Background.....	430
2. Users and Targets.....	431
VI. EVALUATING THE DEMAND HYPOTHESIS.....	433
A. <i>Predictions of the Demand Hypothesis</i>	433
B. <i>Indicator Methodologies and Observed Levels of Accountability</i>	434
1. International Country Risk Guides	434
2. Freedom in the World Rankings	435
3. Minorities at Risk Variables	438
C. <i>Comparing Predictions to Observations</i>	440
D. <i>Grounds for Objections to Each Indicator</i>	442
VII. EXPLAINING THE FAILURES OF THE DEMAND HYPOTHESIS.....	443
A. <i>Users: Not Stakes, But Interest</i>	443
1. Hedging by Users.....	444
2. Insulation from Consequences of Decisions.....	444
3. Unwillingness by Users to Undermine Indicators	445
B. <i>Targets: Not Interest, But Capacity</i>	447
1. Lack of Credibility or Means	447
2. Inability to Marshal Credible Threats of Exit.....	448
3. Barriers to Collective Action	450
C. <i>Indicator Generator Reluctance to Bolster Accountability</i>	454
VIII. THE SUPPLY HYPOTHESIS	455
IX. EVALUATING THE SUPPLY HYPOTHESIS.....	457
A. <i>The User-Dependence and Competitiveness of the Generators</i>	457
1. International Country Risk Guides	457
2. Freedom in the World Rankings	458
3. Minorities at Risk Variables	459
B. <i>Comparing Predictions to Observations</i>	459
X. EXPLAINING THE FAILURES OF THE SUPPLY HYPOTHESIS.....	460
A. <i>The Role of Reputation</i>	460
B. <i>Identity and Participation</i>	461
C. <i>The Professional Ethic</i>	462
XI. CONCLUSION.....	463

I. INTRODUCTION

Decision-making is difficult in a complicated world. As the number of considerations involved in making a decision increases, it becomes harder and harder to gather all the information needed to make an informed choice. Proliferating considerations also introduce uncertainty as to how different factors should be weighed 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. So it should not be surprising that would-be buyers rely on consumer reports, which assemble data on a variety of options within a product class, list them side-by-side, and assign ratings to them. In choosing which college or law school to attend, prospective students often turn to rankings that not only gather in one place relevant information on comparable schools, but also tell students how relatively important each piece of information should be in determining which school to attend. Loan officers lean on credit ratings in choosing whether or not to extend credit to a would-be homebuyer, and school districts point to school funding formulas in apportioning funding for schools.

Measures that summarize complex realities can make decision-making easier. We 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. Indicators are becoming more popular in international settings, from the U.N. 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 simplification of reality,”⁵ that they represent a

1. Willem F.M. de Vries, *Meaningful Measure: Indicators on Progress, Progress on Indicators*, 69 INT’L STAT. REV. 313, 315 (2001) (defining “statistical indicators as single numbers . . . that try to capture a more or less complex reality”).

2. U.N. DEVELOPMENT PROGRAMME (UNDP), *Human Development Reports (HDR)*, available at <http://hdr.undp.org/en/statistics/> [hereinafter UNDP].

3. Richard Cantor & Frank Packer, *Sovereign Credit Ratings*, CURRENT ISSUES IN ECONOMICS AND FINANCE, June 1995, at 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 & Nicolas Fasel, *Quantitative Human Rights Indicators—A Survey of Major Initiatives* (Mar. 2005) (unpublished article, available at <http://www.aboakademi.fi/institut/imr/research/seminars/indicators/Background.doc>) (discussing “major attempts and approaches to develop quantitative human rights and related indicators”).

5. De Vries, *supra* note 1, at 315.

“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 important decisions are made on the basis of indicators, moreover, concerns about the “accountability deficit in the growing exercise of transnational regulatory power”⁸ suddenly apply to indicator generation, as well.

Is it possible to hope that defects in indicators, as well as the process by which they are generated, might be remedied as indicators become more widely used? After all, consumer products that are widely used and subject to competition tend to become more dependable and well-adapted to their intended uses; indicators may similarly become more accurate, and the process by which they are developed marked by greater accountability, as they are employed to make critical decisions. In this Article, I attempt to explain how levels of accountability in the generation of governance indicators are determined. I then reason from this explanation to speculate about how hopeful we should be about accountability in global administration generally.

We can postulate two means by which levels of accountability in indicator generation are determined. The first we can name is the Demand Hypothesis: that the users and targets of indicators demand either accuracy or influence from indicator generators, and that the response of generators to these demands determines observed levels of accountability. The second might be called the Supply Hypothesis: that indicator generators want their indicators to be used, and that they provide higher levels of accountability in order to attract users. In this Article, I will test how well each of these hypotheses explains observed levels of accountability in the generation of three indicators: the International Country Risk Guide (ICRG) ratings produced by the Political Risk Services Group (PRS Group), a for-profit private company;⁹ the Freedom in the World ratings generated by Freedom House, a non-profit non-governmental organization;¹⁰ and the Minorities at Risk (MAR) variables developed by the MAR Project, an

6. AnnJanette Rosga & Margaret L. Satterthwaite, *The Trust in Indicators: Measuring Human Rights*, 27 BERKELEY J. INT’L L. 253, 256 (2009).

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 188, 201 (Thomas B. Jabine & Richard P. Claude eds., U. of Pa. Press 1992).

8. Benedict Kingsbury et al., *The Emergence of Global Administrative Law*, 68 LAW & CONTEMP. PROBS. 15, 16 (Summer/Autumn 2005).

9. PRS GROUP, *International Country Risk Guide (ICRG)*, <http://www.prsgroup.com/ICRG.aspx> (last visited Sept. 24, 2010) [hereinafter PRS Int’l Country Risk Guide].

10. FREEDOMHOUSE.ORG, *Freedom in the World*, <http://www.freedomhouse.org/template.cfm?page=15> [hereinafter Freedom in the World].

academic project at the University of Maryland.¹¹ I will then consider some of the reasons each hypothesis succeeds or fails to describe what is observed, and conclude by consolidating the hypotheses into a single explanatory thesis.

Part II provides an overview of governance indicators: what they are, who generates them, how they are used, and the reasons they are employed. Part III elaborates on some of the terminology that will be used in this Article and addresses some preliminary concerns that must be sorted out before an earnest discussion of accuracy and accountability in governance indicators can be attempted. Part IV presents the Demand Hypothesis, and explains why it seems like a reasonable hypothesis as to how accountability in indicator generation is determined. Part V describes the nature and uses of each of the indicators that will be analyzed here; then, Part VI tests the Demand Hypotheses against the test case indicators by comparing the levels of accountability the Hypothesis would predict in each test case to the levels actually observed. Part VII advances a set of explanations for why the Demand Hypothesis predictions deviate from reality.

Part VIII lays out the Supply Hypothesis, offering it as an alternative to the Demand Hypothesis that may address its shortcomings. Parts IX and X test the Supply Hypothesis against the test case indicators, and, as with the Demand Hypothesis, then attempt to account for any shortfall between its predictions and reality. Finally, in Part X, I conclude by tying the lessons of this analysis together.

II. AN OVERVIEW OF GOVERNANCE INDICATORS

Governance indicators can measure an extremely diverse array of phenomena, and are generated and used by several different types of entities. The proximate causes of the increasing reliance on governance indicators are relatively plain, but the deeper reasons for this shift for quantification are less clear, thus inviting speculation. This section will provide some factual background on governance indicators, and suggest a few reasons for their increasing appeal in aid, investment, journalism, and academia.

A. *What Are Governance Indicators?*

Put simply, governance indicators attempt to measure the quality of governance within a country. Defining “governance” or “good

11. MAR, *The MAR Project*, <http://www.cidcm.umd.edu/mar/> (last visited Oct. 19, 2010).

governance” is difficult, in large part because the term is so capacious.¹² Many definitions attempt to specify the particular components of good governance. Thus, back in 1994, the World Bank defined “good governance” as “epitomized by predictable, open and enlightened policy making; a bureaucracy imbued with a professional ethos; an executive arm of government accountable for its actions; and a strong civil society participating in public affairs; and all behaving under the rule of law.”¹³ Marie Besançon has resorted to even greater detail:

A well governed nation provides: rule of law; political and civil freedoms; medical and health care; schools and educational instruction; roads, railways, the arteries of commerce; communications networks; a money and banking system; a fiscal and institutional context within which citizens can prosper; support for civil society; and a method of regulating the sharing of the environmental commons. Together, the management, supply and delivery of some or most of these goods constitutes governance, and the extent to which nation-states do or do not so perform can—at least in theory—be *measured*.¹⁴

Daniel Kaufmann et al. of the World Bank, in their widely cited World Governance Indicators (WGIs),¹⁵ define governance by means of six dimensions: voice and accountability, political instability and violence, government effectiveness, regulatory burden, rule of law, and control of corruption.¹⁶

The only problem with these definitions of good governance is that they do not necessarily help us identify governance indicators that do not explicitly label themselves as such. We could say that any indicator which measures some aspect of good governance, as described above, should be called a governance indicator—but an easier approach might just be to note that all governance indicators tend to (1) choose some aspect of the relationship between a government and its society, (2)

12. See, e.g., Daniel Kaufmann & Aart Kraay, *On Measuring Governance: Framing Issues for Debate 1* (World Bank Institute, Roundtable on Measuring Governance Issues Paper, 2007), available at http://info.worldbank.org/etools/acportal/docs/Kaufmann%20_Measuring%20_Governance.pdf (“governance writ large is a concept that defies easy definition, and even some of the commonly accepted dimensions of governance, such as democratic accountability, government effectiveness, or rule of law, are themselves subject to definitional ambiguities.”).

13. WORLD BANK, *GOVERNANCE: THE WORLD BANK’S EXPERIENCE* (1994).

14. Marie Besançon, *Good Governance Rankings: The Art of Measurement 1* (World Peace Foundation, WPF Report No. 36, 2003), available at <http://belfercenter.ksg.harvard.edu/files/wpf36governance.pdf> (emphasis in original).

15. STEVEN RADELET, *CHALLENGING FOREIGN AID: A POLICYMAKER’S GUIDE TO THE MILLENNIUM CHALLENGE ACCOUNT 34* (Center for Global Development 2003) (characterizing the WGIs as “the most comprehensive and best quality database available on governance indicators”).

16. Daniel Kaufmann et al., *Governance Matters IV: Governance Indicators for 1996-2004 5* (World Bank Policy Research Working Paper 3630, June 2005).

identify some normative framework within which this relationship can be judged, and (3) measure how well the actual relationship within a country fulfills the normative framework. Consequently, governance indicators do not explore inter-governmental relations—trade negotiations, extradition practices, and representation at international bodies. Likewise, they do not study purely private practices—family life, religious organization, production, and trade—unless such spheres are significantly influenced by government activities.

Essentially all governance indicators, then, attempt to quantify the quality of some aspect of the relationship between a government and its society. Some governance indicators, like those used by U.N. treaty bodies to monitor state compliance,¹⁷ measure human rights protections, while others, like the indices produced by Transparency International,¹⁸ assess perceived levels of corruption. Yet others, such as the Heritage Foundation Index of Economic Freedom, measure economic freedom.¹⁹ Because the points of interaction between a government and its society are innumerable, governance indicators can understandably take an endless variety of forms.

This does not mean, however, that any indicator can be a governance indicator. Excluded from the category, for example, should be the U.N. Human Development Index (HDI), which composes indices measuring life expectancy, educational attainment, and adjusted GDP per capita into a single measure meant to represent levels of human development within a country.²⁰ While the HDI certainly may be correlated with certain aspects of how a government relates to its society, the Index does not purport to directly measure any aspect of the relationship between a government and its society. Similarly, indicators that simply measure economic or financial performance within a country, such as the ICRG Economic Risk Index,²¹ should not really be considered governance indicators, though government policies almost certainly affect their levels.

17. See, e.g., Rosga & Satterthwaite, *supra* note 6, at 257-58 (examining how “the treaty bodies—assisted by an expert institution charged with supporting the treaty bodies, the U.N. Office of the High Commissioner for Human Rights—have begun to use indicators in their monitoring of State compliance”).

18. TRANSPARENCY INTERNATIONAL, GLOBAL CORRUPTION BAROMETER (2009), available at <http://www.transparency.org/content/download/43788/701097/> (last visited Sept. 9, 2010).

19. HERITAGE FOUNDATION, 2010 *Index of Economic Freedom*, <http://www.heritage.org/Index/> (last visited Sept. 9, 2010).

20. UNDP, *supra* note 2; see also de Vries, *supra* note 1, at 322-23 (describing method by which HDI is calculated).

21. PRS GROUP, INTERNATIONAL COUNTRY RISK GUIDE METHODOLOGY 7 (2009), available at <http://www.prsgroup.com/PDFS/icrgmethodology.pdf> (“[t]he overall aim of the Economic Risk Rating is to provide a means of assessing a country’s current economic strengths and weaknesses.”).

B. Who Develops Governance Indicators?

Governance indicators are primarily developed by four different kinds of entities: official aid agencies, non-government organizations (NGOs), academics, and private companies. Examples abound of each kind of indicator generator. Perhaps the most widely respected governance indicators are the WGIs generated by the World Bank.²² The World Bank also produces internal Country Performance and Institutional Assessments (CPIAs)—quantitative measures that guide the allocation of development funds²³—as well as the Business Environment and Enterprise Performance Surveys (BEEPS), which it developed jointly with the European Bank of Reconstruction and Development (EBRD), and which “assess the constraints to private sector growth.”²⁴ Notably, the Asian Development Bank (ADB)²⁵ and African Development Bank (AfDB)²⁶ also follow the CPIA model.

Non-governmental organizations generate an array of governance indicators, from the Corruption Perceptions Index and Index of Economic Freedom produced, respectively, by Transparency International²⁷ and the Heritage Foundation,²⁸ to the array of governance measures tracking competitiveness and government-civil society cooperation generated by the World Economic Forum,²⁹ to the

22. WORLDWIDE GOVERNANCE INDICATORS PROJECT, *Governance Matters 2009*, <http://info.worldbank.org/governance/wgi/index.asp> (last visited Sept. 24, 2010); see also RADELET, *supra* note 15, at 34.

23. INTERNATIONAL DEVELOPMENT ASSOCIATION, *IDA Resource Allocation Index (IRAI)—2008*, <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21359477~menuPK:2626968~pagePK:51236175~piPK:437394~theSitePK:73154,00.html> (last visited Oct. 19, 2010).

24. THE EUROPEAN BANK FOR RECONSTRUCTION AND REDEVELOPMENT, *Economic Surveys*, <http://www.ebrd.com/pages/research/analysis/surveys.shtml> (last visited Sept. 24, 2010).

25. ASIAN DEVELOPMENT BANK, *Performance Based Allocation*, <http://www.adb.org/ADF/PBA/default.asp> (last visited Oct. 19, 2010); see also ASIAN DEVELOPMENT BANK, *REFINING THE PERFORMANCE-BASED ALLOCATION OF ASIAN DEVELOPMENT FUND RESOURCES 1 (2008)*, available at http://www.adb.org/Documents/Policies/ADF/Performance_Based_Allocation/Refining-Performance-Based-Allocation.pdf.

26. AFRICAN DEVELOPMENT BANK GROUP, *2007 COUNTRY PERFORMANCE ASSESSMENT RATINGS 2 (2008)*, available at <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/30735051-EN-BANK-GROUP-CPIA-CPPR-GR-2007-DISCLOSURE.PDF> [hereinafter 2007 COUNTRY PERFORMANCE ASSESSMENT RATINGS].

27. TRANSPARENCY INTERNATIONAL, *supra* note 18.

28. HERITAGE FOUNDATION, *supra* note 19.

29. See, e.g., WORLD ECONOMIC FORUM, *The Africa Competitiveness Report 2009*, <http://www.weforum.org/en/initiatives/gcp/Africa%20Competitiveness%20Report/index.htm> (last visited Sept. 19, 2010) (scoring economic competitiveness of African countries); WORLD ECONOMIC FORUM, *Global Governance Initiative Annual Report 2009*, http://www.weforum.org/pdf/Initiatives/GGI_Report06.pdf (last visited Sept. 19, 2010) (“provid[ing] . . . a

well-publicized Freedom in the World ratings released annually by Freedom House.³⁰ Academics similarly are responsible for a myriad of governance indicators, with some projects pursued by individual academics³¹ and others institutionally supported by universities, such as the University of Maryland³² and National Taiwan University.³³

Finally, some private companies have governance indicators that measure political risk, such as the factors produced by Business Environment Risk Intelligence (BERI)³⁴ and the ICRGs generated by PRS.³⁵ The indicators produced by BERI and PRS are available only for a price.

C. Uses of Governance Indicators

Governance indicators have three predominant uses: determining allocations of development funding, guiding investment flows, and supporting journalism and academic analyses. To begin with the first: governance indicators are sometimes used to guide the allocation of development funding among countries, or to identify a threshold which recipient candidates must cross before they may be eligible for funding. A number of international donors in recent years have begun to demonstrate greater selectivity in administering their loans, based in part on research by World Bank economists Craig Burnside, David Dollar, and Paul Collier indicating that aid is positively correlated with growth in countries with good policies and strong institutions, but

scorecard to measure the effort exerted by governments, international organizations, business and civil society” toward the Millennium Development Goals).

30. FREEDOMHOUSE.ORG, *supra* note 10.

31. See, e.g., Kenneth A. Bollen, *Issues in the Comparative Measurement of Political Democracy*, 45 AM. SOC. REV. 370 (1980) (presenting an index of political democracy); Christopher Clague et. al., *Contract-Intensive Money: Contract Enforcement, Property Rights, and Economic Performance*, 4 J. ECON. GROWTH 185 (1999) (proxying the enforceability of contracts and security of property rights by the ratio of non-currency money to the total money supply).

32. MAR, *supra* note 11 (coding “the status and conflicts of politically-active communal groups”).

33. ASIAN BAROMETER, *Program Overview*, <http://www.asianbarometer.org/newenglish/introduction/ProgramOverview.htm> (last visited Sept. 13, 2010) (surveying “attitudes and orientations toward political regime, democracy, governance, and economic reform” in East Asia).

34. BUSINESS ENVIRONMENT RISK INTELLIGENCE, *Business Risk Service*, <http://www.beri.com/brs.asp> (last visited Sept. 13, 2010) (advertising risk assessment services, including a Political Risk Index).

35. PRS GROUP, *supra* note 21, at 2 (“[t]he aim of the political risk rating is to provide a means of assessing the political stability of the countries covered by ICRG on a comparable basis.”).

exhibits little correlation in countries with poor policies and weak institutions.³⁶

Thus, as already mentioned, the World Bank's International Development Association (IDA) has implemented a Performance-Based Allocation system that steers funds toward countries that demonstrate better "performance in implementing policies that promote economic growth and poverty reduction,"³⁷ as reflected in the internally developed Country Policy and Institutional Assessment (CPIA) ratings.³⁸ Developed internally, and until recently unavailable to the public, the CPIAs assess countries' (a) economic management, (b) structural policies, (c) policies for social inclusion and equity, and (d) public sector management and institutions, thereby determining how IDA resources will be allocated among countries.³⁹ The ADB introduced a system for Performance-Based Allocation in 2001 that uses the same performance criteria as the IDA to distribute aid, though potential aid recipients are rated according to these criteria by the ADB itself, not the IDA.⁴⁰ The AfDB implemented a similar system, likewise featuring internal assessments, in 1999.⁴¹

As for the threshold approach, the main example to date of development agencies using governance indicators to set up thresholds for recipient eligibility is the Millennium Challenge Corporation (MCC), which was established in January 2004⁴² and signed its first

36. Steven V. Hook, *Ideas and Change in U.S. Foreign Aid: Inventing the Millennium Challenge Corporation*, 4 FOREIGN POL'Y ANALYSIS 147, 157-58 (2008) (pointing to studies by Craig Burnside & David Dollar, *Aid, Policies, and Growth*, 90 AM. ECON. REV. 847 (2000), and Paul Collier & David Dollar, *Aid Allocation and Poverty Reduction*, 45 EUR. ECON. REV. 1 (2002) as instrumental in shifting American foreign assistance toward an emphasis on good governance, thereby enabling the creation of the MCC).

37. International Development Association, *How IDA Resources Are Allocated 2010*, <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:20052347~menuPK:2607525~pagePK:51236175~piPK:437394~theSitePK:73154,00.html> (last visited Sept. 13, 2010).

38. See, e.g., INTERNATIONAL DEVELOPMENT ASSOCIATION, COUNTRY POLICY AND INSTITUTIONAL ASSESSMENTS: 2008 ASSESSMENT QUESTIONNAIRE 2-3 (2008), available at <http://siteresources.worldbank.org/IDA/Resources/73153-1181752621336/CPIA2008questionnaire.pdf> (describing the CPIA assessment process and criteria).

39. *Id.*

40. ASIAN DEVELOPMENT BANK, REVIEW OF THE ASIAN DEVELOPMENT BANK'S POLICY ON THE PERFORMANCE-BASED ALLOCATION OF RESOURCES 1, 22-24 (2004), available at http://www.adb.org/Documents/Policies/ADF/Performance_Based_Allocation/performance-based-allocation.pdf.

41. 2007 COUNTRY PERFORMANCE ASSESSMENT RATINGS, *supra* note 26, at 2-3.

42. MILLENNIUM CHALLENGE CORPORATION, *About MCC*, <http://www.mcc.gov/mcc/about/index.shtml> (last visited Sept. 13, 2010) [hereinafter About MCC].

compact in April 2005.⁴³ While there are few other overt examples of aid agencies using threshold or “challenge” funding, the MCC’s approach has the potential, depending on its eventual success, to exert great influence in the development world.⁴⁴

The MCC identifies candidates for assistance based on the requirement that per-capita income levels fall beneath certain thresholds set by the World Bank,⁴⁵ and 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.⁴⁶ 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 indicators found in the Ruling Justly category, in order to become eligible for assistance.⁴⁷ As of January 2010, the MCC had approved \$6.9 billion in compacts with 19 countries.⁴⁸

The MCC’s uniqueness, and hence its high profile, stems from three factors that distinguish it from the other multilateral schemes:⁴⁹ first, its indicators determine absolute eligibility for funding, rather than simply adjusting allocations among recipients; second, its indicators are developed by third parties on the basis of publicly available data, as opposed to being created by agency staff on the basis of confidential determinations; and third, significant amounts of funding are involved, ranging from the \$60 million compact signed with Vanuatu in 2006⁵⁰ to

43. MILLENNIUM CHALLENGE CORPORATION, *Madagascar, U.S. Sign Millennium Challenge Aid Compact* (Apr. 18, 2005), available at <http://www.america.gov/st/washfile-english/2005/April/20050418192454liameruollennoccm0.5341761.html>.

44. See, e.g., Steven Radelet, *Pushing Not Pulling Reforms: Delivering Aid through Challenge Grants*, in *THE NEW PUBLIC FINANCE: RESPONDING TO GLOBAL CHALLENGES* 510, 511 (Inge Kaul & Pedro Conceição eds., United Nations Development Programme 2006) (suggesting that donors are increasingly beginning to differentiate aid strategies by offering “greater national policy ownership, more flexible and attractive aid modalities, and larger, more predictable and longer term resource commitments” to countries that demonstrate good governance). See also Doug Johnson & Tristan Zajonc, *Can Foreign Aid Create An Incentive for Good Governance? Evidence from the Millennium Challenge Corporation* (Apr. 11, 2006) (unpublished manuscript) (on file at http://www.people.fas.harvard.edu/~tzajonc/mcc_wp_apr06.pdf) (finding preliminary evidence that countries respond to MCC incentives by improving their scores on indicators).

45. MILLENNIUM CHALLENGE CORPORATION, *Step 1: Identify Candidate Countries*, <http://www.mcc.gov/pages/selection> (last visited Oct. 19, 2010).

46. MILLENNIUM CHALLENGE CORPORATION, *Selection Indicators*, <http://www.mcc.gov/pages/selection/indicators> (last visited Oct. 19, 2010).

47. MILLENNIUM CHALLENGE CORPORATION, *infra* note 110, at 2.

48. About MCC, *supra* note 42.

49. Radelet, *supra* note 44, at 514.

50. MILLENNIUM CHALLENGE CORPORATION, *Vanuatu*, <http://www.mcc.gov/mcc/countries/vanuatu/index.shtml> (last visited Oct. 19, 2010).

the \$700 million compacts signed with Morocco and Tanzania in 2007⁵¹ and 2008,⁵² respectively. Recipients are given expansive discretion in determining how these funds should be spent.⁵³

Governance indicators are also commonly used in journalism and academia to reflect the state of governance in a country. The most famous indicators—Transparency International’s Corruption Perceptions Index, the Heritage Foundation’s Economic Freedom Index, and Freedom House’s Freedom in the World ratings—are each cited hundreds of times per year in magazines and newspapers.⁵⁴ These citations often include references to particular countries’ scores⁵⁵—so that coverage functions as criticism or approbation of a country’s governance—or, occasionally, assessments of changes in governance within regions or worldwide.⁵⁶ The indicators produced by development agencies and private companies receive far less journalistic coverage.⁵⁷ However, all three types of indicator generators commonly find their products featured in academic publications, most commonly in statistical analyses of how some dependent variable depends on various aspects of governance.⁵⁸

Finally, governance indicators—especially those generated by private companies—influence private investment decisions. According to BERI’s website, “[s]ince 1966 BERI has consistently refined the

51. MILLENNIUM CHALLENGE CORPORATION, *Morocco*, <http://www.mcc.gov/pages/countries/overview/morocco> (last visited Oct. 19, 2010).

52. MILLENNIUM CHALLENGE CORPORATION, *Tanzania*, <http://www.mcc.gov/pages/countries/overview/tanzania> (last visited Oct. 19, 2010).

53. Radelet, *supra* note 44, at 514-15.

54. A LexisNexis search of Major Newspapers and Magazine Stories from Jan. 1, 2008 to Dec. 31, 2008 yields 57 results for “Heritage Foundation” and “Economic Freedom,” 299 results for “Freedom House,” and 447 results for “Transparency International” and “Corruption.”

55. See, e.g., *Call for Vietnam to Adopt Long-Term Strategies, Reforms*, ASIA PULSE NEWS, Dec. 3, 2008 (noting that Vietnam “is currently ranked 120th out of 157 countries in the Heritage Foundation’s Index of Economic Freedom”); *Wife of Sudanese Islamist Leader Accuses Sudanese Officials of Corruption*, SUDAN TRIB., Dec. 31, 2008 (observing that “Sudan has consistently named by watchdog Transparency International as one of the most corrupt countries in Africa”); *Is the Phillippine Press Free?*, MANILA BULL., Oct. 29, 2008 (explaining that Freedom House ranks the Phillippines as “Partly Free”).

56. See, e.g., Gary J. Bass, *Despot Watch*, N.Y. TIMES, Oct. 12, 2008 (noting that “Freedom House’s annual report for 2007 grimly noted that there had been reversals for liberty in one-fifth of the world’s countries”).

57. A LexisNexis search of Major Newspapers and Magazine Stories from Jan. 1, 2008 to Dec. 31, 2008 produces 10 results for “International Country Risk Guide,” 2 results for “Business Environment Risk Intelligence,” 1 result for “World Bank” and “World Governance Indicators,” and 0 results for “World Bank” and “Country Policy and Institutional Assessment.”

58. See, e.g., Jeffrey Sachs et al., *Ending Africa’s Poverty Trap*, 1 BROOKINGS PAPERS ON ECON. ACTIVITY 117, 121 (2004) (exploring the statistical relationship between the Transparency International, WGI, and ICRG governance indicators, and economic growth in tropical sub-Saharan Africa).

system developed by the firm to evaluate risks faced by companies with international operations. Executives benefitting from this experience potentially have a strategic advantage.”⁵⁹ This utility, the private generators argue, leads to significant influence in the business world, with the PRS Group’s website claiming that its products are used by “over 80% of the top companies in the world.”⁶⁰ Christine Arndt and Charles Oman agree, noting that “[d]issatisfaction with the traditional ratings systems has greatly reinforced international investors’ attention to the quality of governance, and their demand for governance indicators, in developing countries.”⁶¹ Indeed, according to Arndt and Oman, “[i]nterviews undertaken for this study of 10 major internationally active banks and companies confirmed . . . the strong recent growth in such investors’ predilection for using governance indicators in their lending.”⁶²

D. Motivations for the Use of Governance Indicators

1. Proximate Motivations

Arndt and Oman identify four developments that have led to an increased interest in governance, and, consequently, greater generation and use of governance indicators. First, the “spectacular growth of international investment in developing countries over the last 15 years” led to “international investors’ major newfound interest in the quality of governance in developing countries,” mostly as a way of managing the risk to which their assets were now exposed.⁶³ Second, the “perceived disappearance of the Communist threat” with the end of the Cold War liberated donors to focus more on the internal governance of their aid recipients, rather than concentrating exclusively on their geopolitical significance.⁶⁴ Third, “growing perceptions . . . of a relative failure or inadequacy of policy reforms widely undertaken in the 1980s and 1990s” led to an increased appreciation of the ways in which “strong markets require good governance.”⁶⁵ Finally, and related to the third point, the New Institutional Economics pioneered by Douglass North and others deepened the understanding of the ways in which institutions

59. BUSINESS ENVIRONMENT RISK INTELLIGENCE, *supra* note 34.

60. THE PRS GROUP, *supra* note 9.

61. CHRISTINE ARNDT & CHARLES OMAN, USES AND ABUSES OF GOVERNANCE INDICATORS 38 (2006), available at http://www.governance.unimaas.nl/training_activities/aau/download/Papers/Usesofabusesofgovernanceindicators%5B1%5D.pdf.

62. *Id.* at 38.

63. *Id.* at 15-16.

64. *Id.* at 16-17.

65. *Id.* at 17-18.

influence and interact with “long-term economic growth, enhancement of human welfare and societal development.”⁶⁶

2. Fundamental Motivations

If the proximate motivations identified above do explain why governance became an increasingly prominent consideration for aid agencies, journalists, academics, and private investors in the first decade of the new millennium, though, they do not necessarily explain why there was a turn to governance *indicators* as a means of ascertaining the quality of governance. A detailed exploration of the sociology of governance quantification is beyond the scope of this Article, but some possible motivations for the move to indicators are suggested by two seminal works of social science: *Seeing Like a State*, by James C. Scott,⁶⁷ and *Trust in Numbers*, by Theodore Porter.⁶⁸

3. Simplification and Uniformity

Scott examines the ways in which modern state elites aspiring to “‘take in charge’ the physical and human resources of the nation and make them more productive”⁶⁹ have often sought to do so by “rationalizing and standardizing what was a social hieroglyph into a legible and administratively more convenient format.”⁷⁰ He identifies a couple of reasons that such an approach has often seemed preferable. Because such elites have often been outsiders to the communities they have sought to understand, control, and improve, they have needed methods of comprehending the salient characteristics of these communities quickly and accurately.⁷¹ This requires, first, a focus on particular features of each community to the exclusion of others, so that the elites obtain a “static and myopic,” if readily comprehensible, view of the community.⁷² Second, elites require the same sorts of information from different communities, so that communities can be compared to one another and their progress charted over time; “the same objective standard” has to be “applied throughout . . . regardless of local context.”⁷³ The consequence of these dual imperatives—toward

66. *Id.*

67. See JAMES C. SCOTT, *SEEING LIKE A STATE: HOW CERTAIN SCHEMES TO IMPROVE THE HUMAN CONDITION HAVE FAILED* (Yale U. Press 1998).

68. See THEODORE PORTER, *TRUST IN NUMBERS: THE PURSUIT OF OBJECTIVITY IN SCIENCE AND PUBLIC LIFE* (Princeton U. Press 1995).

69. SCOTT, *supra* note 67, at 51.

70. *Id.* at 3.

71. *Id.* at 45.

72. *Id.* at 46.

73. *Id.* at 44.

simplification and uniformity—is the production, for elites, of a “systematic and synoptic” view of the communities they seek to manage: a view that is much more useable and accessible to such outsiders than one predicated on intricate local knowledge, which would inevitably involve “judgments that are complex, susceptible to fraud, and easily overtaken by events.”⁷⁴

While governance indicators may not be a technology of control akin to the other instruments of simplification that Scott discusses, their use does seem motivated in large part by the same considerations Scott describes. The interactions between a government and society are obviously intensely complicated, and the most important features of these interactions vary significantly from country to country. Yet aid organizations, journalists, academics, and international investors are all usually outsiders to the countries that they seek to influence, study, or profit from, and each of these actors will often be primarily interested in comparing these countries to one another in allocating aid, approbation, or investment funds, or merely in seeking to develop universally applicable models of how governments and societies interact. Understanding and comparing these countries as outsiders, then, will usually be greatly facilitated by simplified and uniform assessments of these countries, making it likely that accomplishing these dual objectives is one of the main reasons governance indicators are employed.

4. Bolstering Credibility and Concealing Judgment

In *Trust in Numbers*, Porter addresses the increasing turn over the last two centuries to quantification in science, particularly social science, and public policy. To us, living in the first years of the twenty-first century, the use of numbers to communicate information, test relationships, and make decisions may seem natural and obvious, but Porter suggests that numbers have often been employed not because they are the most natural descriptive or decisive mode, but because they confer the appearance of objectivity upon judgments and thus enhance their credibility,⁷⁵ especially in circumstances of high stakes and mistrust.

Porter describes two circumstances that have given rise to quantification that should be of particular interest to us. First: when faced with decisions fraught with serious political consequences,

74. *Id.*

75. PORTER, *supra* note 68, at 7-8 (“[t]he appeal of numbers is especially compelling to bureaucratic officials who lack the mandate of a popular election, or divine right. Arbitrariness and bias are the most usual grounds upon which such officials are criticized. A decision by the numbers has at least the appearance of being fair and impersonal.”).

technocrats (*e.g.*, Army engineers increasingly participating in the selection of water projects in the early twentieth century) have turned to cost-benefit analysis in an attempt “to create a basis for mutual accommodation in a context of suspicion and disagreement,” thereby “promot[ing] procedural regularity” and “giv[ing] public evidence of fairness.”⁷⁶ As Porter explains, such analysis is often subject to “pressures to reify its terms, to deny the validity of human judgment, to lust after the impersonality of purely mechanical objectivity.”⁷⁷ Second, when knowledge communities have lacked cohesion, when they have been bereft of “sharp borders on the outside” and “an effortless shared understanding,”⁷⁸ Porter observes that they commonly have turned to objectivity and mechanical rigor to substitute for the lack of interpersonal understanding and trust.⁷⁹

These twin aims of quantification—to efface the operation of human judgment and overcome deficits of trust, particularly when political stakes are high or a community lacks cohesion—seem particularly applicable to the generation and use of governance indicators. Especially for aid agencies, the funding decisions they make can have serious political consequences; moreover, for such aid agencies, as well as for journalists and academics, any normative claims about what governance within a state should look like will be subject to heavy contestation, given the absence of a shared global community of values. To give their decisions the appearance of objectivity, to obscure the exercise of judgment that underlies them and overcome tendencies toward mistrust, it is natural for these actors to turn to quantification in the form of indicators. Doing so increases their credibility both with their audiences—the political overseers of aid agencies, the readers and peers of journalists and academics, perhaps even the supervisors of investment analysts—and the targets of their judgments and decisions: states and their populations.

III. CONCEPTS AND CONCERNS

Already, this discussion has involved some terms that need further specification, and the analysis to come will involve more such terms. Moreover, some of the bases on which this analysis will rest need to be

76. *Id.* at 149.

77. *Id.* at 187.

78. *Id.* at 227.

79. *Id.* at 223 (identifying mathematics and quantification as “ways of making knowledge more rigid, standardized, and objective,” thereby going “a long way to reduce the need for personal trust,” given the relative rarity of scientific communities connected by “informal, personal knowledge”).

elaborated upon, if not defended, so that skeptics can at least recognize whether they will disagree with the analysis from the outset. This section will aim to explain key concepts and address foundational concerns regarding their use.

A. Indicator Users, Generators, and Targets

In this Article, I will refer repeatedly to indicator users, indicator generators, and indicator targets. The reference to indicator users should be relatively clear; I mean the aid agencies, journalists, academics, investors, and others who use governance indicators to make decisions or substantiate claims. By indicator generators, I mean the entities that produce governance indicators, whether they be aid agencies, nongovernmental organizations, academics or universities, or private companies. By indicator targets, I have in mind those whom governance indicators measure, that is, states and populations.

Essentially, this Article attempts to test two hypotheses about the behavior of these three types of actors, and to explain why the predictions of the hypotheses diverge from reality. Much of this analysis will be predicated upon certain generalizations about the incentives of the users, generators, and targets of governance indicators. I will assume, for much of the analysis, that indicator users want for their indicators to be as accurate as possible, though when we consider the shortcomings of the Demand Hypothesis we will investigate this assumption in more detail. Indicator generators, it is postulated, want to maximize their revenues (especially if they are private companies) and influence, but to do so with as little expenditure of effort as possible.

Indicator targets are somewhat more complicated. One can readily conceive of at least three types of actors within each target—government, business, and civil society—that might have an interest in indicator levels, and it does not appear that all of these actors will necessarily share the same interests. Governments may want to inflate indicator levels to receive greater inflows of development funding and private investment; business may want to deflate indicators to discourage foreign competition, or inflate indicators to secure additional capital; and civil society may seek to ensure indicators are as accurate as possible, so that oppressive governments become subject to international opprobrium. Even these predictions are just stabs in the dark; the actual motivations and interests of groups within country targets may vary widely depending on the circumstances.

Do such heterogeneity and variation matter? As will be described in the presentation of the Demand Hypothesis, actors interested in indicator accuracy should desire enhanced accountability, since transparency, reason-giving, review, participation, and legality allow

them to identify flaws in the indicators, complain about these flaws, and suggest remedies. Other actors, who are interested only in affecting the levels of indicators without regard to their accuracy, should also want accountability, since otherwise they will have no way of exerting the influence they desire. Consequently, despite the heterogeneity of actors within indicator targets, we can still presume a general desire on their part for enhanced accountability. It is on this basis that the analysis will proceed.

B. Accuracy

Throughout this Article, I refer to accuracy in governance indicators—either as something sought by indicator users or by indicator targets. The Demand Hypothesis posits that accountability levels in the generation of governance indicators are determined, to some extent, by demands for accuracy on the part of users and targets of indicators, while the Supply Hypothesis suggests that accountability is determined by the efforts of indicator generators to demonstrate to users that their indicators are accurate. Some may wonder, however, what accuracy in governance indicators even means, and whether it is in fact achievable. Skeptics may suggest that this analysis is irretrievably flawed, given the impossibility of quantifying complex phenomena subject to value-laden judgments.

Governance indicators cannot measure governance itself directly, since it is an abstract concept, or a “construct.” Instead, the indicators must measure some proxy for governance, based on a hypothesis about how this proxy relates to the construct. M.A. Thomas suggests that whether an indicator which is meant to depict a construct “measure[s] what [it] purport[s] to measure”⁸⁰ depends on the answers to three questions: (1) whether the construct represented by the indicator has content validity, that is, “that researchers rigorously define what it is they wish to measure before they set out to measure it, and that the definition have as much in common as possible with the way the construct is typically defined and used[;]”⁸¹ (2) whether the model linking observable variables to the construct is correctly specified;⁸² and (3) whether the “proposed measure has the same relationships with observable variables that the theory predicts the construct itself to have,” or whether the model has construct validity.⁸³

80. M.A. Thomas, *What Do the Worldwide Governance Indicators Measure?* 12 (June 2008) (unpublished manuscript) (on file at http://www.sais-jhu.edu/faculty/melissa_thomas/What%20Do%20the%20Worldwide%20Governance%20Indicators%20Measure.pdf).

81. *Id.* at 18-19.

82. *Id.* at 19.

83. *Id.* at 9.

Skeptics might contend that governance indicators can never be “accurate” under this three-part test, because they could never pass the first prong; governance is too multifarious, polyvalent, and contested a notion for any single definition of governance to have content validity. Even if this is true, though—a question certainly beyond the scope of this Article—the fact remains that governance indicator users, generators, and targets may all nonetheless feel that the concept is amenable to definition, measurement, and verification, and thus may demand accountability in order to affect perceived accuracy (in the case of users and some targets) or the levels at which indicators are set (in the case of some targets), or supply accountability so as to influence perceptions of accuracy (in the case of generators). Whether any definition of governance can actually have content validity, then, is not really relevant to the actions of users, generators, and targets, so long as they believe the construct is capable of definition. In fact, even doubts on the part of these actors about the possibility of governance content validity may not affect the analysis here, since targets that disagree with the effort to define and measure governance should still seek accountability under the Demand Hypothesis in order to make this viewpoint heard, and generators will still attempt to provide accountability under the Supply Hypothesis in order to reassure users that their definitions are valid.

References to accuracy in governance indicators in this Article, then, should not be taken as presumptions that governance can accurately be measured. Rather, this Article is predicated only on the assumption that the perception of accuracy, or the lack thereof, will motivate users, generators, and targets to seek or supply accountability to some extent. The goal of this Article is to understand how these efforts combine to determine accountability levels in the generation of governance indicators.

C. Accountability

If we are to explore the determinants of accountability in indicator generation, we had better start with a good idea of what “accountability” means. In their foundational paper introducing the Global Administrative Law (GAL) Project, Benedict Kingsbury, Nico Krisch, and Richard B. Stewart explain that accountability can be promoted by “ensuring that [administrative bodies] meet adequate standards of transparency, participation, reasoned decision, and legality, and by providing effective review of the rules and decisions they

make.”⁸⁴ These instruments—transparency, participation, reason-giving, review, and legality—are the components of institutional accountability.

Kingsbury et al. supply us with definitions of these terms, borrowed from domestic contexts. They describe “participation” as “the right of affected individuals to have their views and relevant information considered before a decision is taken.”⁸⁵ Transparency involves “access to information” and “exposing administrative decisions and relevant documents to public and peer scrutiny.”⁸⁶ Reason-giving requires the furnishing of “reasons for administrative decisions, including responses to the major arguments made by the parties or commenters.”⁸⁷ Review provides “[a]n entitlement to have a decision . . . affecting one’s rights reviewed by a court or other independent tribunal.”⁸⁸ Finally, legality involves ensuring that “components and agents within [an] order perform their appointed roles and conform to the internal law of the regime.”⁸⁹ Of these definitions, only that of “review” needs to be adjusted for the purposes of this Article; in the global administrative setting, particularly concerning the actions of unofficial entities, we should look for review not by a “court” or “tribunal,” but by any independent arbiter.

D. Global Administrative Law

If indicator-generating institutions could be considered global administrative bodies, then the theoretical framework developed for the GAL project—in particular, the conceptualization of accountability and its components—could be put to use in this Article. More importantly, any useful conclusions arrived at in this Article regarding how accountability levels are determined in indicator generators could later be examined with a view to generalizing them to other global administrative bodies. But are indicator generators global administrative bodies? Kingsbury et al. define

global administrative law as comprising the mechanisms, principles, practices, and supporting social understandings that promote or otherwise affect the accountability of global administrative bodies, in particular by ensuring that they meet adequate standards of transparency, participation, reasoned

84. Kingsbury et al., *supra* note 8, at 17.

85. *Id.* at 37.

86. *Id.* at 38.

87. *Id.* at 39.

88. *Id.*

89. *Id.* at 44.

decision, and legality, and by providing effective review of the rules and decisions they make.⁹⁰

The authors then explain that global administrative bodies include

formal intergovernmental regulatory bodies, informal intergovernmental regulatory networks and coordination arrangements, national regulatory bodies operating with reference to an international intergovernmental regime, hybrid public-private regulatory bodies, and some private regulatory bodies exercising transnational governance functions⁹¹

Indicator-generating institutions certainly seem to possess the broad characteristics of global administrative bodies, at least in the case of governance indicators. Through the governance indicators that they develop, such organizations do “control or supervise [something] by means of rules and regulations,”⁹² and thus act as regulatory bodies, and they do affect matters of “particular public significance”; their standards for assigning scores and ratings to countries both codify and establish norms of state behavior, and the application of these standards to countries leads to judgments by both private actors (*e.g.*, investors) and public actors (*e.g.*, aid agencies) regarding the fitness of other entities (governments, businesses, and civil society) for support.

However, applying the label “global administration” to the profit-seeking activities of a private business, the advocacy efforts of an NGO, or the research of a university project may trouble some observers. If the PRS Group, Freedom House, and the MAR Project are global administrative bodies, one might ask, then why not classify the *New York Times*, whose reporting certainly affects the investment-procuring and aid-receiving fortunes of many countries, as a global administrative body—or even the Travel Channel? The answer cannot be that the PRS Group, Freedom House, and the MAR Project, along with other indicator generators, respect accountability requirements that most media outlets do not, for such a condition would render nugatory one of the main purposes of the GAL project: “to ensure legality, accountability and participation in global administration,”⁹³ as Kingsbury et al. put it. A response that makes more sense is not necessarily that global administrative bodies already possess the virtues that the GAL project attempts to imbue them with, but that such bodies have sufficient normative and practical influence that demands for these

90. *Id.* at 17.

91. *Id.*

92. Oxford English Dictionary, “regulate” (11th ed. Rev. 2008).

93. Kingsbury et al., *supra* note 8, at 55.

virtues make sense. By setting themselves up as impartial arbiters of governance and by allowing—indeed, seeking—a situation in which their judgments act as the basis for public and private actions with significant public repercussions, indicator-generating institutions open themselves up to demands for greater accountability. As a consequence, these institutions are GAL bodies, and analysis of the extent to which they provide accountability becomes appropriate.

IV. THE DEMAND HYPOTHESIS OF ACCOUNTABILITY

According to the Demand Hypothesis, the main types of actors who might be moved by self-interest to demand enhanced accountability are the users and targets of indicators. Because users make decisions based on indicators, these decisions will be less likely to be well-advised if the indicators are inaccurate. Users therefore should have an incentive to push for indicators that are as accurate as possible; because increased accountability will tend to foster improved accuracy, users should demand enhanced levels of accountability. Because the decisions of users will generally affect how the targets of indicators are treated, and because these decisions will often be based to some extent on indicators, such targets should seek greater influence over how indicators levels are set—either to ensure improved accuracy, or just to inflate or deflate their scores. Clearly, the best way to secure this greater influence is through increased accountability. For both users and targets, then, as the stakes of the decisions that depend on governance indicators rise, their demands for increased accountability should grow in volume. All else being equal, this should manifest in increased accountability provision.

A. The Contribution of Accountability

Why does accountability in indicator generation make a unique contribution to accuracy, so that users should be moved by a desire for the latter to demand the former? Each of the GAL components of accountability makes a unique contribution to indicator accuracy. Transparency, or allowing access to information and enabling public scrutiny of decisions, fosters accuracy in indicators by giving interested parties the opportunity to complement their understanding of outcome-oriented deficiencies—indicators failing to accurately depict reality—with procedural deficiencies: indicators being developed through flawed processes. Even if a party interested in improving indicator accuracy is convinced that the indicator is attempting to measure badly a phenomenon that is understood identically by all, the party will be

handicapped in communicating this deficiency to the indicator generator if it cannot examine the generator's methodology and data to understand why the problem is arising. Transparency allows an indicator-generator to harness the efforts of outsiders to diagnose the causes of inaccuracy.

Participation, or the consideration of the views and information of affected parties, allows those interested in the accuracy of indicators to share methodologies and data with indicator generators that they might otherwise neglect. Outsiders may also have suggestions about how to remedy given sources of inaccuracy that do not occur to those inside the organization. Participation thus allows an indicator generator to access the information and ideas of outsiders in an effort to devise treatments for inaccuracy.

Participation does not need to be accompanied by transparency for it to foster accuracy in indicators; even if affected parties do not know how indicators are currently being developed, they can still share useful information and approaches with indicator-creating institutions. But participation is notably enhanced in effectiveness if those who participate have full access to information about how indicators are set. After all, participation becomes more focused and persuasive if participants can funnel information and advice toward perceived weak spots in an institution's procedures.

Reason-giving gives an indicator generator a preemptive opportunity to explain why, appearances to the contrary notwithstanding, any detected inaccuracy is spurious. The requirement of reasoned decision-making also allows outsiders to reappraise initial judgments as to the accuracy of an indicator, or the adequacy of indicator-generating processes and information, by seeing how the processes are applied to information to generate the given indicator—as well as to judge whether decisions are actually being made on the basis of the processes and information revealed by greater transparency. If, after examining reasoned accounts of this application, interested parties still find the information, processes, or indicators wanting, they can point to deficiencies in the reasoning, or to ways in which the reasoning indicates that processes have been poorly applied to information, in the course of either participation in indicator development, or appeal of certain indicators to an arbiter of some sort. Reason-giving thus allows indicator generators to justify the levels at which indicators have been set, to convince outsiders that processes are sound, and to use outsiders to review decisions and processes.

This brings us to review. One can conceive of a system whereby users and targets of indicators, despite opportunities to participate in the creation of indicators, feel strongly that given indicators are incorrect, or even that failures to implement suggested processes or to accept proffered information were unjustified. The opportunity to appeal final

decisions by an indicator generator regarding the levels at which particular indicators should be set, or the processes or information used in setting these levels, allows users and targets of indicators to circumvent institutional myopia, inertia, or indifference and instead plead their case before a disinterested arbiter. From an indicator generator's viewpoint, more importantly, recourse to the judgment of a third party makes it less likely that an outsider will persist in claims that a particular decision or process was unjustified.

Legality, finally, works in tandem with the other accountability components to ensure that their operation is actualized. Without some assurance that individuals within a generator are actually following the processes revealed by transparency, or carrying out the actions agreed upon through participation, or implementing changes suggested by an arbiter, these other components cannot make their full contribution to improving accuracy. Likewise, outsiders must have some assurance that other outsiders are not frustrating their efforts by taking advantage of secret accountability processes.

B. The Provision of Accountability by Generators

Transparency, participation, reason-giving, review, and legality all therefore allow an indicator generator to address inaccuracy more effectively than it could on its own, by harnessing the efforts and knowledge of outsiders to diagnose the sources of inaccuracy, enlarge pools of information and methodologies, and review decision-making. Moreover, by giving outsiders institutionalized methods of getting involved in indicator generation, these accountability components allow generators to control demands for accuracy and to prevent the pitch of these demands from escalating. Reason-giving and review also help to defuse demands for accuracy by justifying the decisions generators have made and by conferring upon these decisions the imprimatur of fairness. If indicator generators choose to respond to demands for accuracy, then, they can do so most effectively and efficiently by ramping up accountability.

The Demand Hypothesis predicts, then, that as the decisions based on indicators involve progressively higher stakes, the intensity of demands by users for accuracy (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.

C. Method of Evaluating the Hypothesis

How does one test this Hypothesis? The method of this Article will be to select three test governance indicators, chosen to represent some variety in the nature of the indicator generator and the predominant users of the indicator. After examining the nature of these users and the types of decisions made based upon the indicators, I will hazard predictions as to how much accountability, under this Hypothesis, should be provided by each indicator generator. These predicted levels of accountability will then be compared to the observed levels, and discrepancies noted. A reexamination of the circumstances and incentives facing the indicator users, generators, and targets will then follow, with the aim of explaining why the reality deviates from the predictions when it does so.

This analysis will largely focus on accountability “on the books”—on the amounts of transparency, participation, review, and reason-giving suggested revealed by publicly available documentation of each indicator. Legality will be considered only where there is some reason to believe a generator is not following its own publicly declared processes. While this approach fails to take into account informal accountability—non-proceduralized, spontaneous responses to information requests, attempts to participate, and complaints about indicator levels—there is a reason for doing so. Aside from the difficulty of actually observing such informal accountability, a good argument can be made that such accountability is less effective than formal, proceduralized accountability, because it depends on outsiders taking the initiative to seek and generate accountability where there appears to be none. It seems fair to the generators, then, to assume that the information available on their websites and in their documentation accurately represents the accountability levels that they furnish.

V. THREE TEST CASES

This Article will take as test cases for its hypotheses three indicators, all of which relate to governance and each of which occupies a different subregion of this field. I will look at (1) the ICRGs 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 MAR indicators developed by the Center for International Development and Conflict Management at the University of Maryland. In this section, I will describe the main features of each indicator-generating institution and the users and targets of each indicator; in Part VI, I will evaluate how well these test cases conform to the predictions of the Demand

Hypothesis.

A. *International Country Risk Guides, PRS Group*

1. Institutional Background

The PRS Group is a for-profit company that has been headquartered in Syracuse, New York since its founding in 1979.⁹⁴ The Group's business largely revolves around two products: PRS ratings, and ICRGs.⁹⁵ Both the PRS and the ICRGs are country risk ratings, but only the ICRGs explicitly measure the present state of governance in the countries analyzed.⁹⁶

In 2008, ICRG produced risk ratings for 140 "countries important to international business."⁹⁷ ICRG country risk 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" and an "overall, or composite, risk rating."⁹⁸ The political risk rating is composed of 12 components, each of which measures a different aspect of political risk and is capped at a different maximum level that varies between indicators; some components are further broken down into subcomponents, each of which is assessed individually.⁹⁹ The components, when added, sum to a maximum of 100 points, and countries are classified as "Very High Risk," "High Risk," "Moderate Risk," "Low Risk," and "Very Low Risk" based on this score.¹⁰⁰

94. PRS GROUP, *About Us*, http://www.prsgroup.com/AboutUs_Overview.aspx (last visited Aug. 26, 2010).

95. *Id.*

96. PRS GROUP, *PRS Methodology*, https://www.prsgroup.com/PRS_Methodology.aspx (last visited Aug. 26, 2010) (PRS country reports 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).

97. PRS Int'l Country Risk Guide, *supra* note 9.

98. PRS GROUP, *ICRG Methodology*, https://www.prsgroup.com/ICRG_Methodology.aspx (last visited Aug. 26, 2010).

99. PRS GROUP, *supra* note 21, at 3-7 (For example, government stability, described as "the government's ability to carry out its declared program(s), and its ability to stay in office," is specified to derive from a combination of government unity, legislative strength, and popular support; each subcomponent has a maximum of 4 points, meaning that government stability as a whole is capped at 12 points). *Id.* at 3.

100. *Id.* at 7.

2. Users and Targets

The PRS Group declares that the ICRG ratings are “[u]sed by institutional investors, banks, multinational corporations, importers, exporters, foreign exchange traders, shipping concerns, and a multitude of others,”¹⁰¹ including “over 80% of the top companies in the world (as ranked by Fortune).”¹⁰² This suggests that the ICRGs have some influence on investment flows that may be large in magnitude.

The PRS Group also offers “sizable discounts on its business products . . . especially for academic use.”¹⁰³ A substantial collection of academics has apparently taken the PRS Group up on this offer, using ICRG ratings as measures of political risk, governance, and corruption in a wide range of papers. Thus, the ICRG ratings have been used to analyze the impact on growth of secure property rights and contract enforcement,¹⁰⁴ the relationship between corruption and public investment,¹⁰⁵ and the effects of aid flows on the quality of governance,¹⁰⁶ to name just a few examples.

Finally, the ICRG ratings are incorporated into other composite governance indexes, the most prominent of which is probably the annually updated WGIs developed by Kaufmann et al. Kaufmann et al. have developed six governance indicators—measuring voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption—by aggregating 340 individual variables drawn from 35 different sources produced by 32 organizations.¹⁰⁷ The ICRG’s political risk ratings play a very important role in these indicators, with at least one ICRG political risk component playing a role in every WGI and 10 of the 12 components included overall.¹⁰⁸ The WGIs are widely

101. *Id.* at 1.

102. PRS GROUP, *FAQ*, <http://www.prsgroup.com/FAQ.aspx> (last visited Sept. 19, 2010).

103. PRS GROUP, *Academic Titles*, <https://www.prsgroup.com/academictitles.aspx> (last visited Sept. 19, 2010).

104. 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* 67 (Christopher Clague ed., Johns Hopkins U. Press 1997).

105. Vito Tanzi & Hamid Davoodi, *Corruption, Public Investment, and Growth*, in *THE WELFARE STATE, PUBLIC INVESTMENT AND GROWTH* 41 (Hirofumi Shibata & Toshihiro Ihuri eds., Springer 1998).

106. Stephen Knack, *Aid Dependence and the Quality of Governance: Cross-Country Empirical Tests*, 68 S. ECON. J. 310 (2001).

107. Daniel Kaufmann et al., *Governance Matters VII: Aggregate and Individual Governance Indicators 1996-2007* 1, 8 (World Bank, Policy Research Working Paper No. 4654, June 2008), available at http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1148386.

108. *Id.* at 72-78 (only religious tensions and socioeconomic conditions are excluded).

respected and influential.¹⁰⁹ Perhaps most notably, the MCC uses several World Bank indicators as selection criteria to determine which countries will receive its development grants.¹¹⁰ Five of the MCC's 17 criteria are derived from the PRS Group's ICRG ratings by way of the World Bank's WGI¹¹¹—including the Control of Corruption indicator, good marks on which is a *sine qua non* for MCC eligibility.¹¹²

The ICRGs' targets are the 140 "countries important to international business" for which political risk ratings are published by the PRS Group.¹¹³ Naturally, the political risk ratings evaluate the performance of national governments, but because the ICRGs are used by investors and development agencies to guide investment and disbursement decisions, the business interests and, indeed, people of these countries are also affected by the levels at which the ICRGs are set, and should be considered targets, as well.

B. *Freedom in the World Ratings, Freedom House*

1. Institutional Background

Freedom House is a non-profit NGO that was founded in 1941, with the encouragement of Franklin D. Roosevelt, to bolster U.S. popular support for involvement in World War II.¹¹⁴ The organization is headquartered in Washington, D.C., but maintains offices around the world.¹¹⁵ In 1972, Freedom House began to publish an annual survey of global political rights and civil liberties: its Freedom in the World rankings.¹¹⁶ This publication rates "every country in the world on a series of indicators basic to freedom," aiming to provide "a comparative view of the global state of freedom."¹¹⁷ Freedom House asserts that its "rigorous research methodology has earned the organization a reputation as the leading source of information on the state of freedom

109. See, e.g., RADELET, *supra* note 15, at 34 ("the most comprehensive and best-quality database available on governance indicators"); ARNDT & OMAN, *supra* note 61, at 49 ("probably the most carefully constructed and widely used indicators").

110. MILLENNIUM CHALLENGE CORPORATION, GUIDE TO THE MCC INDICATORS AND THE SELECTION PROCESS 3-4 (2009), available at <http://www.mcc.gov/documents/mcc-fy-09-guidetotheindicators.pdf> [hereinafter GUIDE TO THE MCC INDICATORS].

111. Kaufmann et al., *supra* note 107, at 72-78.

112. GUIDE TO THE MCC INDICATORS, *supra* note 110, at 2.

113. PRS Int'l Country Risk Guide, *supra* note 9.

114. FREEDOMHOUSE.ORG, *Freedom House: A History*, <http://freedomhouse.org/template.cfm?page=249> (last visited Sept. 26, 2010).

115. FREEDOMHOUSE.ORG, *Staff Directory*, <http://www.freedomhouse.org/template.cfm?page=9> (last visited Sept. 26, 2010).

116. Freedom in the World, *supra* note 10.

117. FREEDOMHOUSE.ORG, *supra* note 114.

worldwide.”¹¹⁸

Freedom House assigns 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.¹¹⁹ In 2009, Freedom House produced political rights and civil liberties ratings for 193 countries and 16 territories.

2. Users and Targets

Freedom House avers that its Freedom in the World 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.”¹²⁰ The ratings certainly appear frequently in the media to support claims about the state of freedom within countries, or to support regional or global generalizations about democratic development. A LexisNexis search of major newspapers and magazines during 2008 yields more than 300 articles referencing Freedom House.¹²¹ A similar search nets fewer than 10 articles mentioning the PRS Group’s ICRGs.¹²²

Like the ICRG ratings, the Freedom in the World ratings are also popular with academics, with papers treating the ratings as proxies for freedom, governance, or democracy, and using them to investigate such phenomena as the effect of democracy on health,¹²³ the relationship between governance and the efficacy of government projects,¹²⁴ and the causal interplay between economic freedom, political freedom, and economic growth.¹²⁵ More often than the ICRG ratings, moreover, the Freedom House indicators are used not just as elements in a regression, but to summarize worldwide or regional changes in freedom, much as the ratings are used in newspapers and magazines.¹²⁶

118. FREEDOMHOUSE.ORG, *Frequently Asked Questions: What is Freedom House?*, <http://www.freedomhouse.org/template.cfm?page=265#1> (last visited Sept. 27, 2010).

119. FREEDOMHOUSE.ORG, *Freedom in the World: 2009 Edition, Introduction*, http://www.freedomhouse.org/template.cfm?page=351&ana_page=352&year=2009 (last visited Sept. 26, 2010).

120. Freedom in the World, *supra* note 10.

121. *See supra* notes 43-45.

122. *See supra* note 46.

123. Álvaro Franco et al., *Effect of Democracy on Health: Ecological Study*, 329 BRIT. MED. J. 1421 (2004).

124. Jonathan Isham et al., *Civil Liberties, Democracy, and the Performance of Government Projects*, 11 WORLD BANK ECON. REV. 219 (1997).

125. W. Ken Farr et al., *Economic Freedom, Political Freedom, and Economic Well-Being: A Causality Analysis*, 18 CATO J. 247 (1998).

126. *See, e.g.*, Fareed Zakaria, *The Rise of Illiberal Democracy*, FOREIGN AFF., Nov.-Dec. 1997, at 22, 23-24 (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

Finally, the Freedom in the World ratings are also used by the Millennium Challenge Corporation as a selection criterion for the disbursement of funds to developing countries, with two of the seventeen MCC threshold indicators furnished directly by Freedom House's political rights and civil liberties ratings.¹²⁷ The World Bank WGI, which incorporates the ICRGs, play a prominent role in determining MCC eligibility, providing five of the twenty-two criteria and the one criterion, Control of Corruption, which every candidate must fulfill to be eligible for funding.¹²⁸ But Freedom House's indicators arguably play a more critical role in MCC selectivity than any of the indicators incorporated into the WGI. First, Freedom House's ratings act as stand-alone threshold indicators; unlike the measures included in the WGI, they are not aggregated with other measures before playing a role in selecting MCC recipients.¹²⁹ Second, the quantized nature of the Freedom House ratings increases the stakes riding on how they are assigned. 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."¹³⁰ As a result, small variations in Freedom House's indicators are likely to have large effects on candidate country eligibility.

C. Minorities at Risk Dataset, MAR Project

1. Institutional Background

The Minorities at Risk (MAR) Project was founded in 1986 by Ted Robert Gurr, then a professor at the University of Colorado,¹³¹ and moved to the University of Maryland along with Gurr in 1988, where it has since been housed in the university's Center for International Development and Conflict Management.¹³² 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

Diamond, *Thinking About Hybrid Regimes*, J. DEMOCRACY, Apr. 2002, at 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").

127. GUIDE TO THE MCC INDICATORS, *supra* note 110, at 2-4.

128. *Id.* at 8-21, 26-28.

129. *Id.* at 3.

130. RADELET, *supra* note 15, at 36.

131. University of Maryland, Ted Robert Gurr, <http://www.bsos.umd.edu/gvpt/gurr/> (last visited Oct. 20, 2010).

132. MAR, About MAR, <http://www.cidcm.umd.edu/mar/about.asp> (last visited Oct. 20, 2010).

do, and what happens to them.”¹³³

MAR data includes both qualitative and quantitative components. The qualitative data comprises a “risk assessment” as to “whether the group is at risk of rebellion, protest, or repression” and a “brief history of the group and its relations with the state[;]”¹³⁴ the quantitative data is composed of approximately 400 variables, including seventy-one identified as “‘core’ variables,”¹³⁵ that code characteristics ranging from “Government repression of group”¹³⁶ to “Group organization and representation”¹³⁷ to “Group concentration.”¹³⁸

The qualitative risk assessment, while “indicator-like” in its assignment of a level of risk to each relevant ethnic group in a given state, is nonetheless better viewed as a non-indicator—no standardized assessment is made of the risk of rebellion, protest, or repression in each situation, but rather all three types of risk are discussed together in a verbose, descriptive manner that does not lend itself to conversion into quantized risk levels.¹³⁹ The quantitative data, on the other hand, takes complex phenomena and reduces them to numbers, and therefore can appropriately be viewed as a set of indicators.¹⁴⁰

2. Users and Targets

The MAR Project states explicitly that it 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.”¹⁴¹ However, the

133. *Id.*

134. MAR, Data, <http://www.cidcm.umd.edu/mar/data.asp> (last visited Oct. 20, 2010) [hereinafter MAR, Data].

135. MINORITIES AT RISK, MINORITIES AT RISK (MAR) CODEBOOK 2-3 (Feb. 2009), available at http://www.cidcm.umd.edu/mar/data/mar_codebook_Feb09.pdf [hereinafter MINORITIES AT RISK].

136. *Id.* at 23.

137. *Id.* at 12.

138. *Id.* at 7.

139. See, e.g., MAR, Data, Assessment for Hazaras in Afghanistan, <http://www.cidcm.umd.edu/mar/assessment.asp?groupId=70001> (last visited Oct. 19, 2010) (noting that “[t]he likelihood of Hazara protest is moderate,” though “[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> (last visited Oct. 19, 2010) (explaining that “Mohajirs [sic] risk for rebellion is moderate,” and that “Mohajirs have a history of persistent protest”).

140. See MINORITIES AT RISK, *supra* note 135, at 1, 6-24 (describing the variables as “quantitative indicators”).

141. CHRISTIAN DAVENPORT, MINORITIES AT RISK, DATASET USERS’ MANUAL 5 (2003), available at http://www.cidcm.umd.edu/mar/margene/mar-codebook_040903.pdf.

most common use of the dataset of quantitative indicators appears to be academic. A tool on the MAR website that searches for publications making use of the dataset between 1993 and 2010 retrieves seventy-one publications between those years,¹⁴² while a 2004 paper by Joanne Manrique finds eighty-five articles referring to the dataset.¹⁴³ These studies sometimes use the MAR data merely to trace the extent of ethnopolitical strife or freedom around the world, but in the larger majority of cases, MAR variables are “operationalized” to enable analysis of the relationship between such variables and other phenomena; Manrique finds the most commonly operationalized variables to be protest and rebellion.¹⁴⁴

In contrast, despite the fact that journalists are an intended audience of the MAR project, journalistic uses for the indicator set seems to be quite limited. A LexisNexis search of major newspapers and magazines from 2008 yields not one non-academic article referring to the Project.¹⁴⁵ The advocacy and public policy uses of the dataset appear similarly restricted. The paucity of journalistic mentions of the project indicates that NGOs are probably not using its data to publicize practices in their own countries, and that states are not employing 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.¹⁴⁶ It is certainly possible, of course, 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. MAR’s website does state that “[m]any are interested in identifying the most threatened and mobilized Minorities at Risk groups,” and consequently offers easily comprehensible maps highlighting “those groups in our database that are engaged in the most mobilization and those that are subject to the highest forms of discrimination and repression for the 2003 time period.”¹⁴⁷ Nonetheless, the most prevalent use of the MAR

142. MAR, Resources, Publications, <http://www.cidcm.umd.edu/mar/publications.asp> (last visited Oct. 19, 2010).

143. Joanne Manrique, *Uses of the Minorities at Risk (MAR) Variables* 3 (Mar. 2004), available at http://www.allacademic.com/meta/p72457_index.html.

144. *Id.*

145. A LexisNexis search of Major Newspapers and Magazine Stories from Jan. 1, 2008 to Dec. 31, 2008 produces only four results for “Minorities at Risk” and (“Gurr” OR “Maryland” OR “CIDCM”).

146. *But see, e.g.*, www.ecoi.net, Iran, Human Rights Issues, Turkmen, <http://www.ecoi.net/189469::iran/328787.321470.8956...mr.321539/turkmen.htm> (last visited Oct. 19, 2010) (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”).

147. MAR, Hot Spots, <http://www.cidcm.umd.edu/mar/hotspots.asp> (last visited Oct. 19,

indicators seems to be academic.

VI. EVALUATING THE DEMAND HYPOTHESIS

What levels of accountability would we predict from our three test indicators under the Demand Hypothesis? If the provision of accountability by indicator generators depends on the intensity of demands for accountability by users and targets of indicators, and if targets and users are likely to make stronger demands if the decisions that they make based on indicators involve higher stakes, we should be able to discern a relationship between the stakes of the decisions made and the accountability observed. This section will explore whether such a relationship is observed.

A. Predictions of the Demand Hypothesis

If the Demand Hypothesis is valid, we should 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, MCC development funding allocations, and academic analyses. Private investment flows are large in magnitude, even for developing countries; as a World Bank report observed in 2005, “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.”¹⁴⁸ MCC grants can amount to hundreds of millions of dollars. Admittedly, academic analyses incorporating the ICRGs involve lower stakes, because they tend to use the Risk Guides to study relationships between governance and other phenomena, rather than to identify governance deficits in particular countries. But the overall stakes involved in decisions based on the ICRGs are certainly high.

The Freedom House rankings apparently play little role in directing investment flows, but they are a critical component of the MCC criteria for allocating funding, and their prominence in journalistic reporting on trends in freedom is notable. Because the reputation of countries can be significantly affected by how they are covered in the media, and because this can affect political influence, prestige within the community of world leaders, and even investment flows (if only indirectly), the stakes involved in the use of the Freedom House ratings

2010).

148. Facundo Martin et al., *The World Bank Group, The Demand for Loans 1* (PUBLIC POLICY FOR THE PRIVATE SECTOR, Apr. 2005), available at <http://rru.worldbank.org/documents/publicpolicyjournal/289martin.pdf>.

are significant, if not so high as those involved in decisions made based on the ICRGs. The MAR variables, on the other hand, incorporated as they are almost exclusively into scholarship, cannot be considered to affect high-stakes decisions.

B. Indicator Methodologies and Observed Levels of Accountability

1. International Country Risk Guides

To test our prediction, we must carefully investigate the methodologies employed in generating our test case indicators, beginning with the ICRGs. The political risk ratings in the ICRGs are subdivided into components and subcomponents; a country is assigned a value for each component or subcomponent by an ICRG analyst, with higher values corresponding to lower risk.¹⁴⁹ 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.¹⁵⁰ The points assigned to each component are then added to yield an aggregate score and a political risk assessment.

ICRG documentation does not describe any systematic provision that has been made for consultation in this process; there is no participatory “step” in the process.¹⁵¹ The PRS Group website does provide contact information that groups wishing to provide input might avail themselves of, but the information pertains to sales contacts; the editorial board, while listed, has no accompanying contact information.¹⁵² There appears to be no provision for third-party review of disputed ratings. 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.¹⁵³ The Group’s provision of the names of the members of its editorial board further bolsters its transparency (though biographies do not accompany these names). But the Group fails make public the

149. The PRS Group, *supra* note 21, at 3-7.

150. The PRS Group, *supra* note 21, at 2.

151. *Id.* at 2-3 (describing the process whereby ICRG ratings generally, and political risk ratings in particular, are determined, and omitting descriptions of any participatory steps).

152. The PRS Group, Contact Us, <http://www.prsgroup.com/ContactUs.aspx> (last visited Oct. 19, 2010).

153. The PRS Group, *supra* note 98

sources of information that it draws on to make its assessments,¹⁵⁴ and does not even explain how these sources of information are chosen, or whether safeguards have been instituted to ensure that its information is accurate, complete, and unbiased. The Group also does not make public the names and qualifications of the staff analysts who do most of the work in calculating ratings.

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.¹⁵⁵ 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”¹⁵⁶; such information could be considered to be either a terse example of reason-giving or an additional source of transparency.

2. Freedom in the World Rankings

Freedom House’s ratings are developed using the services of forty in-house and consultant analysts and seventeen “senior-level” academic advisers.¹⁵⁷ Each country or territory is first assigned to an analyst who is responsible for writing a report describing the situation within each country or territory and for calculating preliminary ratings.¹⁵⁸ In this first step of ratings generation, analysts answer a series of diagnostic questions under the umbrella of either “political rights” or “civil liberties” and assign point values to each response.¹⁵⁹ The political rights rating is based on a checklist of ten questions, each worth a

154. 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.” THE PRS GROUP, *supra* note 21, at 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.

155. *Id.*

156. THE PRS GROUP, *supra* note 21, at 16.

157. Freedom House, Freedom in the World: 2009 Edition, Methodology, http://www.freedomhouse.org/template.cfm?page=351&ana_page=354&year=2009 (last visited Oct. 19, 2010) [hereinafter FreedomHouse.org, Methodology].

158. *Id.*

159. Freedom House, Freedom in the World: 2009 Edition, Checklist Questions and Guidelines, http://www.freedomhouse.org/template.cfm?page=351&ana_page=355&year=2009 (last visited Oct. 19, 2010) [hereinafter FreedomHouse.org, Checklist].

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 fifteen questions that are each worth four points.¹⁶⁰ Freedom House has developed a set of sub-questions to guide the inquiry set up by each diagnostic question,¹⁶¹ though an analyst is not expected to consider every sub-question in assigning a rating to each country.¹⁶² As Freedom House’s website explains, “analysts use[] a 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 in answering the questions.¹⁶³

After each question has been answered and scored, 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.”¹⁶⁴ Questions are answered, and scores assigned, using the points assigned in the previous edition of *Freedom in the World* as a benchmark; changes in raw points are made only “if there has been a real world development during the year that warrants a change.”¹⁶⁵ Moreover, once analysts have answered the diagnostic questions and assigned raw points to countries, Freedom House puts the preliminary ratings that result through a multi-stage review process. 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; and finally, they are scrutinized on a cross-regional basis to ensure comparability and consistency. Any major proposed numerical shifts are subject to intensified investigation in these reviews.¹⁶⁶

The *Freedom in the World* methodology does not suggest that any formal provision is made in the ratings process for participation by outside actors, whether they are institutions making use of the ratings or entities in the countries rated. On the other hand, the “broad range of

160. *Id.* Thus, for example, a country may be awarded as many as 4 points toward its total civil liberties raw score based on the answer to the question, “Is there open and free private discussion?”

161. 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 anti-government conversations to the authorities?” *Id.*

162. *Id.*

163. FreedomHouse.org, Methodology, *supra* note 157.

164. *Id.*

165. *Id.*

166. *Id.*

sources of information” used by analysts to make their determinations does include “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.¹⁶⁷ Freedom House provides lists of staff, analysts, and academic advisers, but contact information for these participants is unavailable,¹⁶⁸ with the Freedom House website providing only general phone numbers and email addresses for each of its offices.¹⁶⁹ Moreover, while ratings pass through several stages of review by staff and consultants working for Freedom House, there do not appear to be institutionalized pathways for outsiders to appeal disputed ratings.¹⁷⁰

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 scores¹⁷¹ to describing the stages of review that initial ratings must pass through.¹⁷² It also provides names and detailed biographies of both the analysts and academic advisers who work on each edition of the rankings, identifying the region to which each was assigned.¹⁷³ Finally, Freedom House makes public a list of nearly 300 publications and broadcasts, and over 150 organizations, which it draws on to produce its ratings, identifying them as “Selected Sources.”¹⁷⁴ Short of opening its deliberations to public scrutiny, then, Freedom House does nearly everything that could be expected to make transparent the process by which it generates ratings.

Similarly, Freedom House provides detailed “country reports” for every state rated in *Freedom in the World*.¹⁷⁵ Each report contains an overview, or a “brief historical background and a description of major recent events, as well as a section summarizing the current state of

167. *Id.*

168. Freedom House, 2009 Edition: Survey Team, http://www.freedomhouse.org/template.cfm?page=351&ana_page=356&year=2009 (last visited Oct. 19, 2010) [hereinafter FreedomHouse.org, Survey Team].

169. Freedom House, Contact Us, <http://www.freedomhouse.org/template.cfm?page=14> (last visited Oct. 20, 2010); FreedomHouse.org, Staff Directory, <http://www.freedomhouse.org/template.cfm?page=9> (last visited Oct. 20, 2010).

170. FreedomHouse.org, Methodology, *supra* note 157.

171. FreedomHouse.org, Checklist, *supra* note 159.

172. FreedomHouse.org, Methodology, *supra* note 157.

173. FreedomHouse.org, Survey Team, *supra* note 168.

174. FreedomHouse.org, 2009 Edition: Selected Sources, http://www.freedomhouse.org/template.cfm?page=351&ana_page=353&year=2009 (last visited Oct. 20, 2010).

175. FreedomHouse.org, 2009 Edition: Country Reports, <http://www.freedomhouse.org/template.cfm?page=21&year=2009> (last visited Oct. 20, 2010).

political rights and civil liberties.”¹⁷⁶ These reports can be viewed either as additional information meant to complement the ratings or as reasoning exercises intended to justify the ratings.

3. Minorities at Risk Variables

MAR develops indicators only for “Minorities at Risk,” or ethnopolitical 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 its self-defined interests.”¹⁷⁷ The project determines which such groups to track according to a set of relatively complex criteria.¹⁷⁸ 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. That is, there are ethnopolitical groups that meet the above criteria and are not included in the dataset.”¹⁷⁹

MAR quantitative dataset includes approximately 400 variables or indicators, grouped into five categories that are themselves further subdivided into sub-categories.¹⁸⁰ The coding is performed primarily by “graduate and undergraduate students who . . . undergo[] a rigorous training period,” with “all coding . . . reviewed by senior editors and by the research director”¹⁸¹ Student coders are guided in their efforts by the labels affixed to the codes that are assigned to each variable,¹⁸² but is unclear whether coders are given access to some more detailed definition of the variables they code, aside from their titles. Some of the information sources relied upon are listed along with the qualitative assessments,¹⁸³ and the MAR documentation explains that “[s]elected hard-copy source materials” are kept in the Project archives, and are

176. FreedomHouse.org, Methodology, *supra* note 157.

177. MINORITIES AT RISK, *supra* note 135, at 1.

178. 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.” *Id.*

179. *Id.* at 2.

180. *Id.* at 5-24.

181. *Id.* at 3.

182. 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. *Id.* at 11.

183. See, e.g., MAR, Data, Assessment for Pashtuns in Afghanistan, <http://www.cidcm.umd.edu/mar/assessment.asp?groupId=70002> (last visited Oct. 19, 2010) (citing reports and articles from the International Crisis Group, New York Times, and Washington Post, among other sources).

available to individual researchers “by arrangement with the project coordinator.”¹⁸⁴ All coding “is conducted using open-source information.”¹⁸⁵

Nonetheless, details on the coding process found in the MAR documentation remain essentially sketchy. The Project does not list the students who have the actual responsibility for converting information describing complex situations into quantitative indicators, so that it is unclear how many such students work on each edition of the indicators, what qualifications the students bring to the process, and whether students specialize in particular countries or regions. Moreover, while some review of student coding apparently occurs, the MAR documentation is ambiguous as to whether this review is systematic and mandatory or selective, multi-stage or single-stage, committee-based or carried out by individual staff.¹⁸⁶ As regards methodology especially, then, the Project fails to provide much transparency. While it might be considered to provide some review in the form of oversight of student coding,¹⁸⁷ there is certainly no review by independent, impartial arbiters at the request of interested parties, just as in the other two cases discussed here.

Unlike the other two indicators studied here, the Minorities at Risk Project makes some provision for participation in the generation of its quantitative variables. Variable values are not reviewed with outside actors before publication, and the Project does not explicitly welcome input about additional sources of information that could be used to supplement the Project’s own data sources,¹⁸⁸ but MAR Project documentation does make it clear that “[r]esearchers are encouraged to carry out their own consistency and validity checks on indicators they use or adapt from the MAR dataset,” adding that “Project staff would greatly appreciate being appraised of the results of such analysis.”¹⁸⁹ Additionally, the MAR website provides lists of project staff and identifies members of its advisory board, accompanying this information with email addresses to facilitate communication.¹⁹⁰ Finally, the Project has reached out to users of its indicators to deal with concerns about selection bias, convening “a workshop at the [American Political Science Association] conference in Philadelphia in September

184. MINORITIES AT RISK, *supra* note 135, at 4.

185. *Id.* at 3.

186. *Id.* at 3-4.

187. *Id.* at 3.

188. *Id.* at 3-4 (describing the process of coding the quantitative indicators, but failing to describe any process of outside review or call for additional sources).

189. *Id.* at 4.

190. MAR, About MAR, Project Staff, <http://www.cidcm.umd.edu/mar/about.asp#staff> (last visited Oct. 19, 2010).

2006 to bring together a group of scholars to help MAR effectively deal with selection bias.”¹⁹¹ As a result, MAR has developed a strategy that it “aim[s] to implement over the coming years, contingent on funding,” 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”¹⁹² Thus, although the MAR Project does not make formal provision for participation in its indicator-generating process, it does both make possible and proactively seek consultation with academic users of the indicators.

Lastly, the Minorities at Risk 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.”¹⁹³ The risk assessment is usually far too terse to justify considering it to be an exercise in reason-giving, with the assessments occupying fewer than five sentences in most cases.¹⁹⁴ The analytic summary, on the other hand, is an approximately 1000-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 ten to fifteen of the most salient indicators.¹⁹⁵

C. Comparing Predictions to Observations

Upon review of these indicator-generating methodologies, how closely does reality adhere to the predictions made by the Demand Hypothesis? To review, the PRS Group provides some levels of transparency in explaining how it arrives at its ratings, if it is less forthcoming about the analysts who produce these ratings and the information they draw on to make their determinations. It provides little opportunity for participation, not even making available the contact information of its directors, and no opportunity for review, though it does provide reasons behind some of its ratings in the form of narrative country reports. The fact that such reports are provided only for twenty

191. MAR, MAR and Selection Bias, <http://www.cidcm.umd.edu/mar/announcement.asp?announcementid=26> (last visited Oct. 19, 2010).

192. *Id.*

193. MAR, Data, *supra* note 134.

194. *See, e.g.*, MAR, Data, Assessment for Roma in Serbia, <http://www.cidcm.umd.edu/mar/assessment.asp?groupId=34506> (last visited Oct. 19, 2010) (offering a six-sentence Risk Assessment).

195. *See, e.g., id.* (including a 929-word Analytic Summary).

to twenty-five of the most important ICRG countries, however, limits their value in justifying the ratings.

Freedom House is a more transparent organization than the PRS Group, providing not only a detailed account of how its ratings are generated, but information about what sources are used in this process and the analysts and experts who participate in ratings development and review. Freedom House does not provide the contact information for these analysts and experts that might allow users and targets to communicate with these individuals and thereby participate in the ratings generation process, but Freedom House's information does come from some sources—NGOs, think tanks, professional contacts, and visits to the region—which would seem to enable outside participation in at least the process of information gathering, if not its analysis. Other opportunities for participation are not made available. While the organization does work with a group of academic experts to review its ratings in a multi-stage process, there is no internal or external review that may be triggered by outside actors. Finally, Freedom House does provide detailed reasons for its ratings in country reports.

The MAR Project, finally, makes impressive provision for participation in indicator generation, soliciting input from the users of its indicators regarding the consistency and validity of the indicators and consulting with academics at, for example, the American Political Science Association's annual conference in order to solicit suggestions on eliminating selection bias. The Project also includes contact information for MAR researchers on the Project's website, giving those interested in offering input into the process access to channels of communication. The Project's marks on transparency are less good; its project documentation explains in summary fashion how its indicators are coded, provides some explication of what each indicator precisely means, and enumerates the sources that it uses in setting its indicators, but details on who actually does the coding, and on the extent of training received by coders, are absent. Enough information is given, then, for outsiders examining the MAR indicators to begin to identify sources of inaccuracy in the way the indicators are developed, but much more information could be provided. Review, either internal or external, is almost entirely absent. The Project does give reasons for a subset of its most important indicators.

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 do not display greater accountability than the Freedom House ratings, or even the MAR variables. The Demand Hypothesis does not, then, seem to do a good job of explaining how indicator accountability levels are set.

D. Grounds for Objections to Each Indicator

This Article does not aim to critique governance indicators, but it is valuable to observe that each of the test case indicators has features which are amenable to criticism. Doing so helps demonstrate that any paucity of demand for accountability cannot result from the design or implementation of these indicators already being beyond reproach.

The ICRGs and the Freedom in the World ratings are both vulnerable to the criticism that their aggregation schemes are almost wholly arbitrary. In measuring governance, why should “Socioeconomic Conditions” be weighted twice as heavily as “Religious Tensions,” or “Investment Profile” be given three times the weight of “Bureaucratic Quality,” as in the ICRG Political Risk Index? Why should “Associational and Organizational Rights” be given the same weight as “Rule of Law,” as in the Freedom House Civil Liberties rating? One answer might be that any weighting of these factors could be subject to criticism, but that still does little to justify the weights chosen.

A related criticism is that there is little reason to believe that the weights chosen should be fixed from country to country. Should “Corruption” command the same weight, as the ICRGs would have it, in Bangladesh as it does in China or Argentina? Should the quality of the “Electoral Process” be considered equally important in judging the levels of governance in Iran and Myanmar, as in the Freedom in the World Political Rights Ratings? It seems likely that some factors will be greater determinants of governance in some countries than in others, and the determination of how much to weight various factors for each country is probably as important as scoring the factors in the first place.

Third, all three test indicators can be criticized for being subject to anchoring effects. Because governance in a given country will be difficult to compare to situations in other countries, the natural tendency for those scoring the target countries will be to compare present governance to past governance, and then reason from past scores to arrive at a present score. Freedom House even formally builds this tendency into its scoring algorithm, changing a country’s score only if a significant development within that country suggests that such a change is necessary. But this means that initial errors in rating country governance will continue to manifest over time, because country targets

will not be given a “clean” score each year and past scores will likely not be revisited.

Finally, while Freedom House states that its rankings are based on travel to the country targets and interviews there with government officials and NGO workers, it is less clear where the PRS Group gets its information, and it does not appear that the MAR Project process involves much direct contact with those living in the target countries. Given that, as Kenneth Bollen explains in the context of political risk indicators, only a “subset of locally reported information . . . travels beyond a country’s borders,” that only a fraction of this information “reaches the United States,” and that “if the filters [that winnow this information] are . . . selective, bias is probable,”¹⁹⁶ any governance assessment that takes place almost wholly from afar is likely to be subject to bias.

A cursory inspection of the indicator-generating methodologies behind the ICRGs, Freedom in the World ratings, and MAR Index, then, suggests a variety of bases on which their accuracy could be criticized. There are grounds for demands for enhanced accountability in indicator generation respecting these three governance indicators; why, then, does the Demand Hypothesis not seem to describe reality very well?

VII. EXPLAINING THE FAILURES OF THE DEMAND HYPOTHESIS

Closer analysis suggests that the Demand Hypothesis has a number of serious flaws that both inhibit demand-making by users and targets and prevent the conversion of these demands into enhanced accountability provisions by generators.

A. *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. The implicit step that is skipped in this relationship is that between stakes and interest; the Demand Hypothesis assumes that when decision-making stakes are high, interest in indicator accuracy should be high, as well. In many contexts, this would not be a bad assumption to make, because actors will tend to scrutinize the bases of their decisions more closely when a lot is riding on these decisions. But in the governance indicator context there are three reasons that the stakes-interest correlation breaks down: (1) investors and agencies are able to engage in hedging behavior when stakes are high, (2) agencies may not be affected by the outcomes of their decisions, and (3) users are

196. Bollen, *supra* note 7, at 199.

unlikely to undermine the indicators they use.

1. Hedging by Users

Even if investors and development agencies have a strong interest in ensuring that they make high-stakes decisions 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. For example, Mikelle A. Calhoun contends “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.”¹⁹⁷ This suggests that investors may not base their investment decisions even in large part on commercial risk ratings, but rather may use these ratings merely to supplement their own in-house analyses.

The MCC, for many of its criteria, relies not directly on measures of corruption or government effectiveness produced by single institutions like the PRS Group, but on aggregate indicators compiled by the World Bank that combine measures from as many as twenty other sources.¹⁹⁸ 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.

2. Insulation from Consequences of Decisions

The hedging argument seems to apply only in part to the MCC, though. If five of the seventeen indicators used by the MCC are furnished by the World Bank Institute in the form of aggregate measures, the other twelve come from single institutions. Even if some of these indicators, like the Immunization Rates criterion provided by the World Health Organization,¹⁹⁹ are sufficiently objective that the MCC might have fewer concerns about accuracy, the same can certainly not be said of the Civil Liberties and Political Rights criteria taken directly from Freedom House’s Freedom in the World ratings.

In this context, though, the lack of consequences for faulty decision-

197. Mikelle A. Calhoun, *Seeking Meaningful Country-Risk Information: A Wealth of Information, a Void in Understanding*, in PERSPECTIVES ON STRATEGIC RISK MANAGEMENT 147, 154 (Torben Juul Andersen ed., CBS Press 2006).

198. GUIDE TO THE MCC INDICATORS, *supra* note 110, at 3-4.

199. *Id.*

making by agencies like the MCC can provide another source of decreased interest in indicator accuracy, even when decision-making stakes are high. Oversight mechanisms do 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,²⁰⁰ if these legislators have a strong interest in ensuring that funds go to well-governed countries, and if they can ascertain how efficiently funds are being allocated along these lines, then those working for the MCC should develop a substantial interest in ensuring that the indicators used to direct allocations accurately reflect reality.

But neither of the conditions specified are necessarily met. Legislators may want aid to be spent effectively, but they also likely want aid to be used in support of American geopolitical aims, and likely also want a certain amount of aid money to be disbursed simply to demonstrate, to constituents and other nations, a commitment to development. Moreover, legislators will have great difficulty directly observing 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 have difficulty monitoring how well such funds are distributed. Finally, the use by the MCC of externally developed measures like the Freedom in the World ratings actually allows the organization to scapegoat entities like Freedom House for any dearth of results that is actually observed by legislators, allowing the Corporation to avoid budgetary consequences. For all these reasons, the MCC, and other development agencies that might rely on indicators to make funding decisions, should not necessarily have a significant interest in the accuracy of these indicators, but rather an interest in justifying expenditures to those who exercise oversight. The two are not necessarily equivalent.²⁰¹

3. 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

200. See, e.g., Steven Radelet, *Will the Millennium Challenge Corporation Be Different?*, 26 WASH. Q., 183 (2003) (describing Congress’s role in allocating MCA funds).

201. This should not be taken to suggest that professionals working in agencies like the MCC are not sincerely motivated to achieve positive development results. Certainly, most of those who enter the development field likely do so because they want to “do good.” But the discipline of consequences for poor decisions can create an interest in successful decision-making that sincerity may not be able to match. We should have little reason to believe that this discipline will be present in organizations like the MCC.

improved accuracy or accountability because such demands risk discrediting measures to which they have linked their own reputations. Investors, by publicizing the inadequacies of political risk services like the PRS Group, risk alerting backers to the weakness of their own strategies. Development agencies similarly court sanction from legislative oversight committees and protests from candidates for aid, and academics open themselves to the charge that their work is invalid. While the users of most consumer products—cars, televisions, software—have little disincentive to protest when such products fail to meet expectations, the users of governance indicators should generally seek either to lodge any criticisms of accuracy or accountability levels in private, or not to make these criticisms at all.

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 almost half of the organization's leadership have spent a substantial portion of their careers working for the federal government.²⁰² 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 U.S. 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.

Such selectively and, to some extent, secretly elevated levels of transparency, participation, reason-giving, and review bolster accountability to some extent, but they also undermine accountability by reducing the legality of indicator generators. When the stated methods by which indicators are developed are subverted by the informal access of a few actors, the foundation on which the other accountability components rest becomes shaky. Other actors can no longer be certain that the indicator generator is actually relying on the methods and information that its statements would indicate, or that their own participation is not being frustrated by the unobservable participation of others, or that the reasons being given for particular decisions are not being supplied by outside interests. Legality is thus a critical, though less conspicuous, component of accountability; if users are able to secure higher accountability levels for themselves at the cost of overall legality, it is not clear that accountability is, on the whole,

202. See Freedom House, Board of Trustees, <http://www.freedomhouse.org/template.cfm?page=10> (last visited Oct. 19, 2010) (providing a list of the organization's trustees that links to biographies of each; according to their biographies, of the 41 trustees listed, at least 18 have had significant experience in government).

bolstered.

B. Targets: Not Interest, But Capacity

The defect in the Demand Hypothesis when it comes to predicting users' behavior, then, appears to be that even if the interests of such actors in indicator accuracy determine the intensity of demands for accuracy and accountability, this is nonetheless poorly proxied by the size of the stakes involved in the decisions made by users. The Hypothesis encounters a different defect when it attempts to predict the behavior of targets. Even if targets have strong interests in ensuring that indicators are accurate, or at least in seeking opportunities to influence how indicators are set—with these interests determined by what rides on the indicators—they may lack the capacity to influence the behavior of indicator generators or users. At least three reasons for this incapacity can be identified: (1) targets often lack the credibility or means needed to make persuasive demands; (2) the inability to marshal a credible threat of exit from the indicator-target relationship undermines the weight of potential demands by targets; and (3) collective action problems reduce the likelihood of target action.

1. Lack of Credibility or Means

Targets, when deciding whether to make demands upon indicator generators for improved accuracy or accountability, almost certainly take into account the chances that their demands will be heeded. Since the making of demands may be costly in terms of time, human capital, and deflection of lobbying resources from other activities, targets may avoid lodging such demands even if they have a strong interest in ensuring that indicators are accurate, or in influencing the way these indicators are set. We can identify a few factors that might discourage targets from engaging in demands.

To begin, those who use indicators to judge the quality of governance in a country likely do so in part because the indicators are generated by ostensibly impartial observers, in contrast to data supplied either by the users' own governments or the businesses and governments of target countries. Consequently, indicator generators will be reluctant to heed the demands of targets for fear of the damage such interactions may cause to their reputations. Unless targets can marshal truly compelling evidence to suggest that an indicator is flawed, indicator generators should be inclined to ignore them—especially if the targets represent countries that have reputations for oppression and secrecy.

Target countries also may 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 conclusions drawn by an indicator generator. Moreover, the education levels of the country's elite, and the sources of education, will play a critical role in determining whether attempts to persuade indicator generators to change their operations will be successful. If a country has few persons that have been educated abroad, and 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. Thus, less wealthy countries will be less able to engage in the persuasion involved in lodging effective demands.

2. 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. This relationship between exit and voice is detailed by Albert O. Hirschman in *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*.²⁰³ Hirschman's model does not have auspicious implications for the success of demands made by targets upon indicator generators, which suggests in turn that targets should be reluctant to expend the resources needed to make these demands.

Hirschman explains that when an organization fails to meet performance expectations, customers or members have two routes through which they can alert management as to its failings: the exit option, in which “[s]ome customers stop buying the firm's products or some members leave the organization,” and the voice option, whereby “[t]he firm's customers or the organization's members express their dissatisfaction directly to management . . . or through general protest.”²⁰⁴ When faced with alternative B to current choice A, a customer or member will resort to the more costly voice option rather than exit if “A's original margin of superiority over B was wide enough to make it worthwhile for him to forego a B that is superior right here and now” in the hopes that voice will lead A to regain its original quality.²⁰⁵ Thus, viewing opportunities for exit as proportional to the

203. See ALBERT O. HIRSCHMAN, *EXIT, VOICE, AND LOYALTY: RESPONSES TO DECLINE IN FIRMS, ORGANIZATIONS, AND STATES* (Harvard U. Press 1970).

204. *Id.* at 4.

205. *Id.* at 38-39.

quality of B, “the role of voice would increase as the opportunities for exit decline, up to the point where, with exit wholly unavailable, voice must carry the entire burden of alerting management to its failings.”²⁰⁶ If exit is easy—if B is equivalent to A—then customers or members are unlikely to expend the effort that voice requires. However, given that “[o]ne important way of bringing influence to bear on an organization is to threaten exit,” “voice is not only handicapped when exit is possible, but also, though in a quite different way, when it is not.”²⁰⁷ 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.”²⁰⁸

But how can Hirschman’s concept of “exit” be of use in analyzing the behavior of indicator targets? Unlike indicator users, the targets of indicators cannot simply choose to stop being measured; at best, it seems, they could refuse to cooperate with the generators of institutions, denying researchers visas and declining to respond to any requests for information. But if targets are motivated by the desire to influence how indicators are set, such measures seem unlikely to accomplish the given aims.

Indicator targets do have another alternative available to them, though—they can attempt to induce the users of indicators to stop using indicators to evaluate the target, so that the target would have effectively “exited” from its status as an object of measurement. Of course, the measurement and indicating activities would continue, but their significance for the target would be decreased. Targets can accomplish this effective exit in a few ways—by pointing out to users the inadequacies of certain indicators, by supporting the use of alternative indicators, or by providing users with supplementary information and analysis that make indicators less useful and hence less likely to be employed. To the extent that generators rely upon users to supply them with revenues, reputation, or influence, the prospect of effective exit by targets and its concomitant, actual exit by users, should frighten generators into taking targets seriously.

How successful is this threat of “effective exit” likely to be in backstopping demands made upon generators by targets? The first two methods described above have the potential to backfire upon targets because of the credibility problem already described in the preceding section. Complaints from a target perceived to be corrupt or repressive about a given indicator may actually bolster its reputation for accuracy;

206. *Id.* at 34.

207. *Id.* at 55.

208. *Id.* at 83.

efforts by such targets to induce users to switch from one indicator to another might foster perceptions that the favored indicators are not impartial and accurate, but actually in thrall to the targets. The first two approaches also include many of the disadvantages that characterize the “voice” option and the absence of which usually makes the “exit” option 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 a costly one, particularly if such users have publicly committed to rely on a given set of indicators. 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, and the MCC cannot simply switch procedures abruptly. Finally, users of indicators will often not have the patience to consider supplementary information required by the third approach, and may not be willing to credit this information even if they are willing to consider it. In particular, the MCC, journalists, academics, and investors likely use indicators mostly because their quantitative nature provides such users with the uniform, “synoptic” view already discussed with reference to James C. Scott’s work.²⁰⁹ It is unlikely that the provision to such users of detailed information of variable scope and quality will be welcomed.

The powerlessness of targets to exit from their relationship with indicator generators, even in an “effective” sense, makes their demands far less powerful. Unless targets, especially in the form of government agencies or civil society groups, derive some benefit from the act of criticizing indicators itself, targets should tend to concentrate their persuasive capital on other, more promising endeavors.

3. Barriers to Collective Action

In 1965, Mancur Olson published *The Logic of Collective Action: Public Goods and the Theory of Groups*,²¹⁰ seeking to explain when “groups of individuals with common interests” will “attempt to further those common interests.”²¹¹ While Olson frequently refers to “organizations,” his work is equally applicable to any groups sharing a common interest, so that it can help explain when targets will act to

209. See *supra* Part II.D.2.a, discussion.

210. MANCUR OLSON, *THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS* (Harvard U. Press 1971).

211. *Id.* at 1.

secure enhanced accountability from indicator generators. Generally speaking, such action seems unlikely.

According to Olson, “[t]he individual member of the typical large organization” occupies a position in which “his own efforts will not 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”²¹² Consequently, large organizations will not be able to “support themselves without providing some sanction, or some attraction distinct from the public good itself, that will lead individuals to help bear the burdens of maintaining the organization.”²¹³ 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”²¹⁴ Such “larger” members have an incentive to pay the cost of providing for some amount of the collective good, even if they must do so alone. Thus, while even in the smallest groups “the collective good will not ordinarily be provided on an optimal scale,” we should expect that “the larger the group, the farther it will fall short of providing an optimal amount of a collective good.”²¹⁵ Moreover, “[i]n small groups with common interests there is . . . a surprising tendency for the ‘exploitation’ of the great by the small.”²¹⁶

It is not immediately evident that enhanced accuracy or accountability represents a public good for indicator targets, however, so that Olson’s model need not necessarily apply here. After all, a target country could conceivably intervene to change accuracy or accountability with respect to *its indicators alone*, through the direct provision of supplementary information or efforts to secure more accountable indicator-generating processes. But information provision is made possible only by high levels of participation and review, and as explained, high levels of these accountability components are really only rendered meaningful if they are accompanied by transparency, reason-giving, and legality, as well. One way or another, a target will thus have to push for greater accountability if it wants to influence how indicators are set.

An indicator generator might conceivably respond to lobbying for increased accountability by providing such enhanced accountability *only* to the target doing the lobbying, just as I surmised above that generators may sometimes provide selectively elevated accountability to some users. If generators prize their reputation for independence and

212. *Id.* at 16.

213. *Id.* at 15-16.

214. *Id.* at 33-34.

215. *Id.* at 34-35 (emphasis removed).

216. *Id.* at 35 (emphasis in the original).

impartiality, though, engaging in such apparent collusion with only some indicator targets will be to their significant detriment. Thus, efforts by targets to secure more accountable indicator-generating processes should, when it succeeds, more usually result in broad changes by the generator than in changes that affect only the particular indicator that the target country demander has in mind. As a consequence, from the perspective of targets, we are dealing here with a public good.

Based on Olson's model, then, we should anticipate that when indicator targets have the capacity individually to secure enhanced accuracy or accountability with respect to the indicators that affect them, and when the welfare of these actors is significantly affected by decisions made based on indicators, demands for accuracy or accountability should become more likely. It is less probable that this capacity threshold will be crossed for a large group of "small" targets than for a small group of "large" targets, though if a large group includes some "large" actors we might still see such demands.

How might we characterize the number, interests, and "size" of governance indicator targets? For all three of the indicators considered here, the number of targets should be large. The ICRGs include political risk ratings for 140 countries, Freedom House evaluates over 200 countries and territories in its Freedom in the World ratings, and the MAR project codes variables on ethnopolitical groups in nearly 120 countries. If each country that is the object of indication is viewed not monolithically as a single target, but as composed of businesses, civil society groups, government agencies, and ethnic groups, each with a potential interest in how the indicators are set, the number of target actors for each indicator becomes quite large indeed.

Each indicator will likely be characterized by a different level of target interest in how the indicator is set. It is hard to imagine that targets are excessively exercised about the levels of the MAR variables, given the almost exclusive use of these variables in academic scholarship that affects the welfare of targets in extremely indirect fashion, at best. But the governments, business interests, and civil society of indicated countries should all have a strong interest in the ICRGs if the ratings are actually used to make investment decisions, even if they are only employed as a check against internal analyses; the lack of transparency regarding exactly how these ratings are used by investors might actually increase target interest in how they are set. The ICRGs' inclusion in the WGI that furnish several of the criteria for the MCC should bolster target interest in indicator levels, especially on the part of government actors within the target countries. However, it should be noted that country targets with larger economies might be less interested in ICRG levels than small-economy targets. For particularly

important investment destinations, it seems likely that investors will rely much less on political risk ratings and much more on their own analyses, displaying the hedging behavior already described. Moreover, the several-hundred-million-dollar development grants through the MCC are likely to make up only a negligible part of the budgets of large-economy countries.

The Freedom House ratings present a somewhat similar picture. The main uses of the ratings are in academia, MCC funding, and journalism. Targets are likely to be little interested in the first use. The second use mirrors that of the ICRGs, except that the Freedom in the World ratings are arguably more decisive in determining MCC allocations than the Risk Guides, since the Freedom House ratings are not aggregated with other indicators to furnish a single criterion and Freedom House ratings tend to be clustered around the median score, which is the cut-off for country candidate eligibility. However, large-economy countries should be much less interested in the role Freedom House ratings play in determining MCC funding than small-economy countries. The same cannot necessarily be said of the journalistic uses of the Freedom in the World ratings; even if journalism does not influence investment decisions or aid allocations in a determinate manner, it does affect a country's reputation and geopolitical clout, and the membership of its rulers in the fraternity of world leaders. Large countries should be as interested in these goods as small countries. However, journalistic coverage of larger countries that incorporates the Freedom House ratings is more likely to be diluted by a mass of other coverage, which should reduce the ratings' importance. Generally speaking, then, the same large-economy and small-economy gap in interest that we anticipated in the ICRG case should appear with respect to the Freedom House ratings.

Finally, we must evaluate the "size" of governance indicator targets. Some target countries are wealthier or represent larger economies. They should thus have more resources to draw on in lobbying efforts, so that if capacity to bring about changes in accuracy or accountability levels in indicators depends on the resources available for lobbying, several targets will be substantially "larger" than the rest. However, it is possible that there are enough countries with large economies and significant resources at their disposal that they might free ride on each other's efforts to improve accuracy and accountability. Thus, if there are a large number of "large" targets, none of the targets may end up taking action. On the other hand, to the extent that targets view themselves as not necessarily sharing the same aims, the group of targets may subdivide into blocs—developing countries, African countries, commodity-producing countries—within which only a few large actors might be found. This would reduce the likelihood of free-riding.

Taking all this together, what are the prospects for action by targets to secure the public goods of enhanced accountability in indicator generation? Even if the group of targets subdivides into blocs, the number of actors within each bloc will be large. However, many of the actors will have significant interests in influencing how some indicators are set, given the importance of the ICRGs and Freedom in the World ratings in shaping investment and development decisions, and within each bloc there should only be a few well-resourced, “large” actors. This might seem to indicate the target action should be favored. However, a serious dilemma faces target countries seeking to influence indicator generation: those countries with the *greatest* capacity to prevail upon the indicator generators to change the way that they set indicators will also be the countries with the *least* interest in ICRG accuracy. 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 the MCC, and (3) be the subjects of a great deal of journalistic coverage, entirely apart from that which relies on Freedom House ratings. The countries that have the most to lose from unfavorable indicators will also be the countries least capable of convincing the generators to change their ways.

Altogether, then, we should expect few demands from indicator targets upon generators for improved accuracy or accountability, even if some targets have a strong interest in the levels at which these indicators are set. A lack of credibility and means to make demands, the absence of the credible threats of exit needed to give force to demands, and collective action problems should all push targets toward silence.

C. Indicator Generator Reluctance to Bolster Accountability

If users are unlikely to make demands for enhanced accuracy because they will often lack the interest, and if targets are unlikely to make such demands because they will usually lack the capacity, another factor complicates the hypothesized relationship between demands and accountability: even when demands for accuracy are made upon generators, the generators may not respond by boosting accountability. In our exposition of the Demand Hypothesis, we postulated that because improved accountability has a unique potential to bolster accuracy and stave off future demands, generators should respond to demands for accuracy by increasing accountability. However, it is actually just as likely that bolstering accountability will tend to intensify, not diminish, demands by outsiders. Because the effects of increased accountability are difficult to pin down, indicator generators may choose to respond to

demands for accuracy by implementing changes in the ways indicators are generated, without bringing outsiders into this process.

On the one hand, ratcheting up accountability seems likely to mitigate external demands upon an indicator generator. After all, providing increased opportunities for participation and review should decrease other demands for improved accuracy, since outsiders will usually prefer to make use of institutionalized channels for making demands. Indicator generators might prefer such an arrangement not only because it lessens the probability that demands will be publicly expressed and hence discredit the indicator, but also because having mechanisms for dealing with these demands reduces their potential to distract the entire organization from the task of indicator generation. 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 extremely 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, reason-giving, and legality. However, it certainly seems unlikely that such improvements in accountability will result merely from demands by outsiders for greater accuracy or accountability, and generators might even prefer to avoid increases in participation and review, given their uncertain effects.

VIII. THE SUPPLY HYPOTHESIS

The Demand Hypothesis is likely incorrect in presuming that users and targets will attempt to secure enhanced accountability from indicator generators and that indicators will be eager to furnish such accountability. Nonetheless, we should still expect users, when deciding

which indicator to employ, to seek out the most accurate indicator. Once an indicator has been chosen, some users may be limited in their capacity to exit to alternative indicators; development agencies like the MCC may have publicly committed to use the indicator in allocating funds, and investors may have built decision-making structures that have a certain indicator, with its idiosyncratic features, at their heart. When a user is still deciding which indicator to employ, it is likely to be drawn to indicators that appear more accurate, all else being equal. This expectation about user behavior forms the foundation of the Supply Hypothesis.

Users will often be handicapped in directly evaluating the accuracy of an indicator. Just as legislators, in attempting to exercise oversight over development agencies, will usually have difficulty identifying whether development outcomes are a consequence of poor funding allocation on the part of the agencies, indicator users will often be hard-pressed to tell whether an indicator correctly measures the phenomena it is supposed to reflect, or whether the results of their own investment or development decisions are a consequence of inaccuracy in the indicators used or other factors. Users tend to judge indicator accuracy not by looking at the indicators themselves, but by looking for markers that tend to 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 further the task of convincing potential users that its indicator has the marks of accuracy. In particular, generators should provide heightened levels of transparency and reason-giving when they seek to persuade users to employ their indicators, so that users may be impressed by the thoroughness of indicator-generating processes and the validity of the generator's indicator-setting approach. 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 under-provided. Moreover, indicator generators will tend to provide greater transparency and reason-giving when potential users could easily do without a particular indicator, and when the use of that indicator by a user will tend to confer a substantial benefit—whether in the form of revenues or influence—upon the indicator generator.

IX. EVALUATING THE SUPPLY HYPOTHESIS

Just as with the Demand Hypothesis, we can evaluate the Supply Hypothesis by seeing what predictions it makes about our test case indicators, and comparing those predictions to the levels of accountability we observe. To begin, we should attempt 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.

A. *The User-Dependence and Competitiveness of the Generators*

1. International Country Risk Guides

The PRS Group depends on revenue from the sale of licenses to its products.²¹⁷ Because the main purchasers of these licenses are private investors, the Group can be said to rely substantially on its relationship with these investors; if investors stop using the ICRGs in their decision-making, PRS Group revenues will suffer. 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 in which they are incorporated) would not directly affect the Group's bottom line, since the MCC's use of the ICRGs does not lead directly to increased revenues for the Group. But academics and the MCC both 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 available, moreover. Many governance indicators exist; Marie Besançon of the World Peace Foundation lists almost fifty.²¹⁸ Some of these indicators measure only particular aspects of governance (such as Transparency International's Corruption Perceptions Index²¹⁹) or cover only particular regions (such as Asian Barometer²²⁰), while others are more comprehensive. There is at least one other privately produced political risk indicator—the Business Environment Risk Intelligence (BERI) ratings²²¹—and several

217. The PRS Group, *supra* note 94.

218. Besançon, *supra* note 14, at 11-34.

219. TRANSPARENCY INTERNATIONAL, *supra* note 18.

220. Asian Barometer, *supra* note 33.

221. Business Environment Risk Intelligence, *supra* note 34.

other indicators measure the quality of the business environment.²²² For either development agency or international investor users of ICRGs, internal analysis of governance and political risk is always an option; as discussed, the World Bank uses internal assessments rather than the WGI published by their own World Bank Institute to guide their Performance Based Allocation of funds,²²³ and many investors base decisions mostly on their own assessments, using commercial risk services like the ICRGs only in supplementary fashion.²²⁴ Thus, private investors, academics, and development agencies have other governance measures that they could choose in case the ICRGs appear insufficiently attractive.

2. Freedom in the World Rankings

In contrast to the PRS Group, Freedom House is not an organization that depends directly upon any of its users for financial support, because it is primarily supported by grants from “various private foundations and government agencies.”²²⁵ Given that the U.S. government supplies 75% of the Freedom House budget,²²⁶ it might be true that the organization would see its funding affected if its ratings proved to be of little use to U.S. development agencies like the MCC. However, given that Freedom House was receiving U.S. 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, though, Freedom House is an advocacy NGO that likely measures its success, and justifies its existence to funders, in part based 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, both financially and in a more existential sense.

As with the ICRGs, moreover, Freedom House does have many competitors in the field of governance indicators, though some of these indicators may not measure the same precise 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

222. See, e.g., *supra* Part II.B, discussion.

223. See *id.*

224. See *supra* VII.A.1, discussion.

225. Freedom House, About Us, <http://www.freedomhouse.org/template.cfm?page=2> (last visited Oct. 19, 2010).

226. See FREEDOM HOUSE, 2005 FREEDOM HOUSE ANNUAL REPORT 24 (stating that of \$26.8 million in total revenues and support received in Fiscal Year 2005, \$20.2 million came from federal grants), available at http://www.freedomhouse.org/uploads/special_report/41.pdf.

default to use of Freedom House's ratings even in the absence of transparency or reason-giving that helps persuade them that the ratings are reliable, basing their choice not on probable accuracy but on reputation or the need for extensive data. I will return to this point in a moment, when we consider the defects of the Supply Hypothesis.

3. Minorities at Risk Variables

Finally, the MAR Project does not derive financial support from its users; the MAR variables are free.²²⁷ 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, though, might be simply to expand the amount of available knowledge about ethnopolitical groups, with only limited concern about how this knowledge is used.²²⁸ The funders whose grants make the MAR Project's activities possible may share this predominant interest in generating new knowledge for knowledge's sake. 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.

B. 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.

What do we see in reality? The PRS Group does provide some levels of transparency in explaining its methodology, though it provides little insight about the analysts who develop ratings and about the information these analysts use. Reports that explain the ICRGs are provided for twenty to twenty-five of the most important countries that are rated, but there is very little scope for participation or review in

227. See MAR, Data, <http://www.cidcm.umd.edu/mar/data.asp> (offering free downloads of qualitative and quantitative data).

228. It is certainly beyond dispute that many academics are more interested in the hypothetical usefulness of their work than the actual usefulness, if they are interested in usefulness at all.

ICRG generation. Freedom House is much more transparent, explaining in detail how its ratings are produced, who participates in this process, and what sources of information are used. Freedom House also provides detailed justificatory reports about every one of the countries and territories that it rates. From its documentation, it is possible to infer that Freedom House also provides some limited opportunity for participation—listing among its sources of information NGOs, think tanks, professional contacts, and visits to the region—and review, though there is no review by neutral arbiters upon the request of outside parties. As for the MAR Project, it is somewhat lacking in transparency, with only limited information available about its ratings methodology, and reason-giving, with reasons given only for a subset of its variables. The Project, like Freedom House, provides no external review, though it does provide a great deal of scope for participation by making the contact information of its staff available, by soliciting feedback from users, and by working with users at academic conferences to improve its methodology.

In the main, then, the predictions of the Supply Hypothesis are borne out by reality. The PRS Group and Freedom House both provide significant 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 some explanatory power. By considering where this Hypothesis might fall short, we can construct a more complete explanation.

X. EXPLAINING THE FAILURES OF THE SUPPLY HYPOTHESIS

A. The Role of Reputation

Transparency and reason-giving are not the only tools available to a generator seeking to entice potential users. The accuracy and reliability of an indicator may be ascertained 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, then 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 well-known but frequently pilloried, then generators may have to work more

to convey the impression of reliability.

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, even though the Supply Hypothesis would predict otherwise. The Freedom House ratings, while extremely well-known, are also often subject to charges from journalists and academics that they display a right-wing bias and are heavily influenced by the political aims of the U.S. government.²²⁹ The ICRGs, though widely used, are much more rarely the object of any critical coverage. The PRS Group may be more secure relying on its reputation to attract users than Freedom House, then.

B. *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 our discussion of the defects in the Demand Hypothesis, moreover, I reasoned that generators should be disinclined to furnish high levels of participation or review under most circumstances, because ratcheting up levels of these accountability components tends only to prompt more demands and criticisms from outsiders. These two propositions still leave room for generators to allow participation and review under some circumstances, though. When generators are motivated not primarily by the need to persuade possible users to rely on their indicators, when it is not too costly to bring a class of outsiders into the indicator development process, and when the outsider-generator relationship is marked by a certain level of trust, generators may increase the levels of participation and review that are available exclusively to that class of outsiders. In short, when indicator generators and a certain type of outside actor share an attribute that frequently brings them into contact, 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 did not have to worry about demands from their

229. See, e.g., Right Web, Freedom House, <http://rightweb.irc-online.org/profile/1476.html> (last visited Oct. 19, 2010) (“[w]hile touting itself as having a ‘bipartisan character,’ Freedom House is often associated with hawkish and neoconservative factions within both major U.S. parties, a fact made clear by many of its current and past supporters and board members.”); Bollen, *supra* note 7, at 205 (noting that “researchers claim that [Freedom House’s] conservative slant has led to incorrect rights and liberties ratings for some countries”).

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, and the problems of credibility and means that likely afflict indicator targets in their efforts to communicate with generators would not have been present, either. In fact, shared identity can explain more in the MAR context than just the observed levels of participation. The shared profession of indicator generators and users may also explain whatever transparency deficits are observed here, and 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, and 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 external, independent arbitrator.

When we first considered the possibility that generators may selectively extend higher levels of accountability to some users, in the context of the Freedom House ratings, I suggested that this prospect raised concerns about legality that may outweigh any other accountability benefits that might result. Should we feel any differently about the selective provision of accountability to the MAR users? I would argue that we should, for two reasons. Selectively elevated levels of accountability may appear unfair, 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, the overall accountability profile of the indicator should not be considered to be injured. The generator is still providing for voice and choice by those who would exercise it.

C. The Professional Ethic

Finally, readers may have already remarked to themselves on the bloodlessness of the Demand and Supply Hypotheses and their accompanying commentary. Humans do not act only when others compel them to do so by their demands, or when it is easy to do so. In our discussion so far, little mention has so far been made of the

professionals who work at indicator-generating institutions themselves. It seems possible, though, that the most powerful determinant of how much accountability is furnished by a generator is not the competitiveness of the environment facing the generator, its dependence on its users, or even the commonality of identity between generator and user, but rather the organizational ethos that characterizes the generator. Open, consultative organizations with intrinsically motivated staff may have such an interest in producing high-quality indicators that they provide high levels of transparency, participation, reason-giving, review, and legality of their own volition, conscious that such openness will tend to foster greater accuracy.

This is certainly an important possibility. Two brief comments should be made regarding it, however. First, it would be exceedingly difficult to incorporate “openness” or “sincerity” into a model of how accountability levels in indicator generation are determined. Unless crude characterizations were employed—for-profit companies are insincere, NGOs are open, multilateral institutions are bureaucratic, and so on—evaluations of how open or sincere a generator is will inevitably depend largely on observed levels of accountability, which is the phenomenon we are trying to explain. Trying to base a model on sincerity thus risks allowing such a variable to spread to explain all the observed variation, making the effort to develop a model an essentially useless exercise. Secondly, it is not clear that sincerity plays as large a role as its proponents might claim. After all, none of the indicator generators analyzed here provide significant levels of review, which would be predicted by the Supply Hypothesis but not by an emphasis on sincerity, unless it be assumed that every generator considered here is marked by a lack of commitment to indicator accuracy. When an organization is trying to do the best it can to improve its influence or revenues in light of competition and scarce resources, sincerity may actually have very little to do with final levels of accountability; it may simply be squeezed out by other considerations.

XI. CONCLUSION

In this Article, I have sought to explain how levels of accountability in indicator generation are determined. I began with a hypothesis that, while intuitively appealing, proved not to explain observed levels of accountability in our three test cases: that when the stakes of decisions made based on governance indicators are high, targets and users of these indicators will demand enhanced accuracy and accountability from generators, which will lead these generators to provide higher levels of accountability. The Demand Hypothesis appeared to fail for a

few reasons. First, even when users make high-stakes decisions based on indicators, they may nonetheless not display serious levels of interest in the accuracy of these indicators. Second, even when targets should be interested in the levels at which indicators are set, they may not have the capacity to act to secure this public good. Finally, even if demands for accuracy or accountability are made by users and targets, generators should still often have good reason to respond to these demands not by improving accountability but by tinkering with their methodologies.

Having lost faith in the Demand Hypothesis, we turned to an alternative: the Supply Hypothesis. I postulated that indicator generators may provide higher accountability levels not because they are demanded by users or targets, but because certain types of accountability help to “sell” indicators to potential users. This Hypothesis encountered greater success when brought into contact with reality, though it still needed to be modified by certain provisos. Rather than levels of competition between indicators and indicator dependence on users determining all of the observed variation in accountability, a role likely needed to be left for reputation and identity.

What are we left with, then? Indicator generators should display higher levels of transparency and reason-giving when they are in competition with other indicators, or when users can do without indicators altogether, and when the 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 any other kinds of actors) share an identity attribute that decreases the costs of furnishing such accountability for both types of actors.

Why is this interesting? The study of governance indicator generation actually does have implications for other global administrative law bodies. The decisive failure of the Demand Hypothesis indicates that we should not expect high levels of accountability to result just because a GAL body engages in consequential behavior; collective action problems, an inability to marshal a credible threat of exit, and the capacity to hedge or duck responsibility on the part of actors affected by the behavior may nonetheless mean that accountability levels remain low. When a GAL body needs to secure compliance or cooperation from other actors, we should expect levels of transparency and reason-giving to be higher, so as to support the persuasive activities of the body. But participation and review should usually be low unless the GAL body administrators and other actors share a community or identity that makes provision of these

accountability components more low-cost and less likely to spiral out of control.

Or, at least, one can hypothesize that these relationships and trends hold true in the larger GAL context. The real test, of course, should lie in comparison with reality.

