

**Politics by Other Means:**

**Economic Expertise, Power, and Global Development Finance Reform**

Jigar D. Bhatt

Submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy  
under the Executive Committee of the Graduate School of  
Arts and Sciences

COLUMBIA UNIVERSITY

2018

© 2018  
Jigar D. Bhatt  
All rights reserved

## ABSTRACT

### **Politics by Other Means:**

### **Economic Expertise, Power, and Global Development Finance Reform**

Jigar D. Bhatt

This dissertation investigates how economic expertise influences development governance by examining how state economists establish methods for decision-making in global development finance. It contributes to debates over expert power by taking a science studies approach to address two problems in existing theories and accounts of experts. First, social reformers, heterodox planning theorists, and development critics from both the left and the right treat rationality and politics asymmetrically. When experts fail, politics has triumphed. When experts succeed, the credit goes to rationality, not politics. Second, within this asymmetrical approach, investigations and explanations of expert power neglect a principal conduit of expert influence: their methods. This dissertation turns the focus to economists' efforts to establish their methods as governing rationales and the effects these methods engender. Doing so allows us to approach particular forms of state rationality such as neoliberalism or managerialism not as processes of depoliticization, of intellectual rationality prevailing over political interests and values, but as explicit political accomplishments with both the power to bring about political effects and the susceptibility to being challenged.

State economists' efforts to establish three paradigmatic development economic methods in particular—governance indicators, growth diagnostics, and randomized controlled trials—and these methods' effects on power relations, decision-making, and the distribution of resources were assessed using an embedded case study design of their use for decision-making in

administering a new development finance fund, the United States Millennium Challenge Account. A mixed methods approach using interviews, documents, and various datasets found that economists could not realize the power of their intellectual rationality without exercising power thought to be the reserve of politicians. Economists had to employ various strategies of power both to gain autonomy from bureaucratic authorities and overcome opposition from expert groups holding alternative rationalities. This involved enrolling bystanders and opponents in their entrepreneurial efforts to establish methods. The more opposition economists faced, the more power they had to exercise and allies they had to enroll. Once enrollment was successful, economists' status was elevated and their methods became indispensable to particular decision-making processes. These new ways of making decisions introduced different biases that elevated economists' concerns, objectives, and ways of knowing. They also impacted the distribution of development finance in ways that exacerbated inequality in at least the short to medium term.

This dissertation's focus on economists' political work and methods has implications for planning practice because it opens up new political possibilities. Rather than treating state expertise and public participation as antagonistic, zero-sum confrontations, planners can pursue democratic values by both "opening up the state" and "getting inside" methods. If orthodox economists had to overcome opposition from groups of opposing experts with competing rationalities then other experts can likewise use political strategies to establish their methods as governing rationales. Even in situations where this is not possible or desirable, understanding methods' political effects can instigate reflective practice and possible change.

## Table of Contents

List of Tables .....	ii
List of Figures .....	v
List of Acronyms .....	viii
Acknowledgements .....	xi
Chapter 1: A Science Studies Approach to Expertise in the State .....	1
Chapter 2: Major Debates over the Power of Experts .....	22
Chapter 3: Research Approach and Methodology .....	47
Chapter 4: The Power of Governance Indicators in Global Development Policy .....	70
Chapter 5: The Power of Growth Diagnostics in Development Planning .....	144
Chapter 6: The Power of Randomized Controlled Trials in Development Projects .....	213
Chapter 7: The Power of State Expertise .....	281
Chapter 8: Economic Expertise, Development Governance, and Planning Practice .....	323
Works Cited .....	341
Appendix A: List of Interviewees .....	377
Appendix B: Tables and Figures for Governance Indicators Chapter .....	380
Appendix C: Tables and Figures for Growth Diagnostics Chapter .....	382
Appendix D: Tables and Figures for Randomized Controlled Trials Chapter .....	389

## List of Tables

Table 3.1: Sample Frame, Target Interviews, and Number of Respondents .....	60
Table 3.2: Secondary Data Sources, Including Dataset Sources .....	61
Table 3.3: Public Events Attended .....	62
Table 4.1: Original MCA Qualification Criteria .....	110
Table 4.2: Central Asian Countries Disqualified from MCA Consideration Due to WGI Scores .....	125
Table 4.3: ADF Performance-Based Shares by Country Group, 2002–2008 .....	139
Table 5.1: Basic Profile of First Nineteen Compacts, 2005–2009 .....	161
Table 5.2: Sector Distribution of First Nineteen Compacts .....	162
Table 5.3: Distribution of Analyses by Constraint in CA Final Reports .....	192
Table 5.4: Number of Times Constraint Identified in Final Report and Classified as Binding .....	193
Table 5.5: Distribution of Identified Constraints Matched to Final Compact Investments .....	195
Table 5.6: Summary Statistics for Compacts Developed with and without Constraints Analyses .....	201
Table 5.7: Distribution of Sectors Among Compacts Developed with Constraints Analyses .....	202
Table 6.1: MCC's Evaluation Record as of 2015 .....	236
Table 6.2: Distribution of MCC Development Funding and RCTs by Sector and Year .....	238
Table 6.3: Sectoral Distribution of RCTs in 3ie Database of Impact Evaluations .....	239
Table 6.4: RCTs Among Agriculture and Rural Development Projects in 3ie's Database .....	244
Table 6.5: MCC's Completed FTT Projects with an RCT .....	249

Table 6.6: Eligibility Criteria for Participation in MCC FTT Projects .....	255
Table 6.7: Key Characteristics Among Early and Late Treatment Groups: MCC-FTT Projects .....	262
Table 6.8: Differences-in-Differences Matrix .....	263
Table 7.1: Sites of Conflicting Rationalities .....	292
Table 7.2: Growth Diagnostics as Obligatory Passage Points .....	305
Table 7.3: Governance Indicators as Obligatory Passage Points .....	307
Table 7.4: Randomized Controlled Trials as Obligatory Passage Points .....	309
Table 7.5: Influential Decision-makers Before and After Stabilization of Methods .....	316
Table B-1: Differences in Income Between Low-income Countries Based on MCA Qualification Criteria Performance, 2004–2012 .....	380
Table B-2: Differences in Income Between Lower Middle-income Countries Based on MCA Qualification Criteria Performance, 2006–2012 .....	381
Table C-1: Distribution of Categorized Data Sources in MCC Constraints Analyses .....	382
Table C-2: Distribution of Categorized Data Sources in MCC Constraints Analyses for High Cost of Finance Constraints .....	384
Table C-3: Distribution of Categorized Data Sources in MCC Constraints Analyses for Market Failure Constraints .....	385
Table C-4: Distribution of Categorized Data Sources in MCC Constraints Analyses for Government Failure Constraints .....	386
Table C-5: Distribution of Categorized Data Sources in MCC Constraints Analyses for Low Social Return Constraints .....	387
Table C-6: Summary Statistics on Analysis Types Found in MCC Constraints Analyses .....	388

Table D-1: List of Completed and Ongoing RCTs at MCC as of Fiscal Year 2014–2015 ..... 389



## List of Figures

Figure 4.1: The Puzzle of Global Disequilibrium According to the US Treasury .....	84
Figure 4.2: Organizational Structure of Bush’s National Security Council .....	92
Figure 4.3: The NSC Interagency Participants and Policy Making Process .....	94
Figure 4.4: Organizational Chart of Key Bush Administration White House and Departmental Personnel Involved in the MCA’s Creation .....	96
Figure 4.5: Indicator Work Plan Circulated to the MCA Interagency Group Members .....	111
Figure 4.6: Plot of Countries' WGI Government Effectiveness Point Estimates with Confidence Intervals and Quartile Distributions .....	116
Figure 4.7: Individual Components of Treasury's Proposed Composite Governance Index .....	118
Figure 4.8: Stylized Fact Representing Positive Relationship between WGI's Rule of Law Indicator and Per Capita Income .....	120
Figure 4.9: Example of Analysis Conducted by Treasury Economists in the Indicator Working Group .....	120
Figure 4.10: Country Coverage Comparisons Between WGI and Treasury's Proposed Indicator .....	122
Figure 4.11: Differences in GNI Per Capita among Low-Income Countries Qualifying or Failing to Qualify for the MCA .....	133
Figure 4.12: Differences in GNI Per Capita among Low-Income Countries Passing or Failing the WGI .....	134
Figure 4.13: Differences in GNI Per Capita among Lower Middle-Income Countries Qualifying or Failing to Qualify for the MCA .....	135

Figure 4.14: Differences in GNI Per Capita among Lower Middle-Income Countries	
Passing or Failing the WGI .....	136
Figure 4.15: Scatterplot Comparison of Countries' 2000 and 2009 WGI Voice and	
Accountability Estimates .....	141
Figure 5.1: Growth Diagnostic Decision Tree .....	149
Figure 5.2: MCC Examples of Growth Constraints.....	151
Figure 5.3: Early Compact Development Process .....	153
Figure 5.4: Simplified Model of an Aid Negotiation .....	158
Figure 5.5: Example of Test 1: Shadow Prices for Transport as a Binding Constraint .....	171
Figure 5.6: Example of Test 2: The Relationship Between Investment Levels and	
Corruption .....	172
Figure 5.7: Test 3: Examples of Coping Mechanisms .....	173
Figure 5.8: Test 4: A Search for "Camels and Hippos" .....	174
Figure 5.9: Compact Development Process Charts .....	189
Figure 5.10: Distribution of Funding Across Sectors Between Compact Cohorts .....	203
Figure 5.11: Percent of Total Compact Funds Allocated to Urban or Rural Activities .....	207
Figure 5.12: Percent of Funds Going to Reform Activities Before and After CAs Affected	
Compact Design (2010) .....	209
Figure 6.1: The Rise of RCTs in Development .....	217
Figure 6.2: The RCT Network .....	225
Figure 6.3: MCC's Matrix Management Structure, 2004–2007 .....	230
Figure 6.4: RCTs and Project Implementation .....	247
Figure 6.5: Implementation Structure of FTT Projects .....	250

Figure 6.6: Example of a Simplified Program Logic Framework from the El Salvador	
FTT .....	251
Figure 6.7: Randomized Roll-out Design for FTT Projects .....	252
Figure 6.8: Schema of Training Approach With and Without a Randomized Roll-out	
Evaluation .....	267
Figure 6.9: Training Roll-out in Ghana's FTT .....	269

## List of Acronyms

ACDI	Agricultural Cooperative Development International
ADB	Asian Development Bank
ADF	Asian Development Bank's Asian Development Fund
ARD	Agriculture and Rural Development
CA	Constraints Analysis
CEA	Council of Economic Advisors
CEO	Chief Executive Officer
CEP	Coalition for Evidence-Based Policy
CGD	Center for Global Development
CPIA	Country Policy and Institutional Assessment
DAC	Development Assistance Committee
DECRG	Development Economics Research Group
ERR	Economic Rate of Return
ESA	Environmental and Social Assessment
FBO	Farmer-Based Organization
FOIA	Freedom of Information Act
FTT	Farmer Technical Assistance and Training
G8	Group of Eight
GDP	Gross Domestic Product
GFCF	Gross Fixed Capital Formation
GNI	Gross National Income
HRV	Hausmann, Rodrik, and Velasco

IADB	Inter-American Development Bank
IFDC	International Fertilizer Development Center
IDA	World Bank International Development Association
IMF	International Monetary Fund
ISSER	Institute for Statistical, Social, and Economic Research
J-PAL	Abdul-Latif Jameel Poverty Action Lab
MCA	Millennium Challenge Account
MCC	Millennium Challenge Corporation
MD	Managing Director
MiDA	Millennium Development Authority
MIT	Massachusetts Institute of Technology
MPR	Mathematica Policy Research
M&E	Monitoring and Evaluation
NCE	Neoclassical Economics
NEC	National Economic Council
NGO	Non-governmental Organization
NIE	New Institutional Economics
NSC	National Security Council
ODA	Official Development Assistance
OECD	Organization for Economic Cooperation and Development
OMB	Office of Management and Budget
PRS	Poverty Reduction Strategy
PRSP	Poverty Reduction Strategy Paper

RCD	Resident Country Director
RCT	Randomized Controlled Trial
SECAL	Sector Adjustment Loan
SUTVA	Stable Unit Treatment Value Assumption
UC	University of California
UN	United Nations
UNDP	United Nations Development Program
USAID	United States Agency for International Development
USG	United States Government
VOCA	Volunteers in Overseas Cooperative Assistance
VP	Vice President
WGI	Worldwide Governance Indicators

## Acknowledgements

Like a child, it takes a village to raise a dissertation. Many kind souls have helped carry this project over the finish line and I owe them all a huge thank you and debt of gratitude. I would like to start by thanking my committee. My advisor, Robert Beauregard, took an earnest interest in this project from its inception. One could not have asked for a better advisor or doctoral program director. Sanjay Reddy has been both a mentor and kindred intellectual spirit. Our conversations have impacted my thinking well beyond the dissertation. Gil Eyal provided insightful input at critical junctures and welcomed me to the wonderful world of expertise studies. Elliott Sclar's commitment to heterodox economic thought and positive change helped keep my focus and Lance Freeman's commitment to rigor kept me honest.

A number of other scholars at Columbia and beyond have been vital to this dissertation. I thank Smita Srinivas for helping me question my assumptions and her invitations to collaborate. Elizabeth Popp Berman was always willing to share insights and her passion for studying economists is infectious. Daniel Immerwahr and Michele Alacevich provided unique perspectives and Moshe Adler always reminded me of the importance of this work. I would like to thank the members of Columbia's 2014 Science, Knowledge, and Technology Workshop and participants in the Initiative for Policy Dialogue's 2015 Advanced Graduate Workshop, especially Akbar Noman, Hemanshu Kumar and Varun Piplani.

This dissertation would not have been possible without the help of family and friends. I could not have finished it without the love and patience of my wife Jennifer Huynh who never stopped believing in the project or me. I could not have begun it without the support and encouragement of my sister Rupal Bhatt and brother-in-law Frank Weichmann. My niece and

nephew, Nisha Singh and Vishal Jani, provided good humor and camaraderie throughout the process. Several friends at Columbia made the doctorate experience memorable. My time at Columbia and since would not have been the same without the friendship of Rose Hackman. I give many thanks to Eric Goldwyn and Lauren Ames Fischer for their friendship and great conversations. I also thank Banke Oyeyinka, John West, Padmini Biswas, and the cohort with which I began this journey: Xiaohong Pan, Sophonie Joseph, and Alexis Perrotta. I appreciate Leigh Brown who was always ready to lend a helping hand. Kun Leng, Mark Choi, and the members of Columbia TRACT were a source of inspiration and hope. Beyond campus, Seth Pipkin was always there when I needed a shoulder (or great mind) to lean on. Karthik Raghavan helped me feel at home at the New School and Dan Robinson helped make writing away from New York easier.

Many friends from my days at MCC have come along on this journey with me. Thanks go to Richard Gaeta and David Ameyaw for their wisdom and early encouragement to pursue a doctorate. Big ups to my “MCC family”—Mark Horowitz, Nasra Ismail, Laura Rudert, and Bikram Sachdeva—who always lifted my spirits. I thank Sarah Rose and Eric Lundgren for always being supportive and our insightful discussions. Farrah Barrios kept me grounded and Emily Andrews and Shiro Gnanaselvam were always in my corner.

Last but not least, I would like to thank the many individuals and key informants who responded to interview requests and spent hours answering my many questions. They prefer to remain anonymous, but without their time, input, and insights this dissertation would not have been possible.



## Chapter 1: A Science Studies Approach to Expertise in the State

### **Introduction**

On April 9, 2014 the Center for Global Development (CGD), a leading development think tank in Washington, DC, hosted a debate at CGD headquarters and online on William Easterly's book, *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor*, published that same year (Easterly, 2014). Easterly, a well-known development economist, debated Owen Barder, a CGD Vice President and Senior Fellow, who was described as a development "thinker and doer", on the following resolution: "On balance, development experts, with their technocratic approach and their indifference to the abuse of political power and of people's rights, have harmed rather than helped the interests of poor people" (CGD, 2014). The audience was asked to vote on whether they agreed or disagreed with the resolution before and after the debate. After one and a half hours of discussion, only two percent of viewers changed their minds (CGD, 2014).

Easterly's hyperbole, the way experts and politics are framed in the resolution, and that so few viewers were moved by the discussion are indicative of an impoverished debate over the role and influence of experts, particularly economists, in global development. This dissertation aims to provide a partial corrective to how the role of experts and their influence in planning, policy, and development are framed and discussed.

This chapter introduces the dissertation by describing why experts and politics are treated as distinct in the planning and development literature. It then presents the dissertation's purpose and the research questions driving its empirical work. The research design and methodology supporting the execution of this empirical work is described next. The chapter then provides a background to the case being investigated by situating it within broader trends in global

development finance and development economics. Afterwards it explains the study's conceptual framework and contributions to research and practice. The chapter concludes with an overview of the remaining chapters in the dissertation.

### **Background to the Research Problem**

This dissertation examines a paradox. Critical theorists from both the left and the right critique the power of experts in planning and administration. Meanwhile, social reformers bemoan that state experts are routinely ignored or dismissed by decision-makers. And, experts themselves decry their lack of influence. Why, then, if critics believe experts are so powerful, do they seem inconsequential by others? This dissertation addresses this paradox by asking whether state experts are powerful, and if so, how, by investigating a key conduit through which expert influence is channeled: the methods that link knowledge with action. My specific focus is on economic methods for decision-making in global development finance.

Two major intellectual limitations have contributed to the paradox. The first is social reformers' linear rational model (Baum, 1980; Weiss, 1979). The second is neo-marxist and postdevelopmental theorists' arguments that state expertise generates a depoliticized "post-political" condition (Beveridge & Koch, 2017; Sachs, 1992). These groups treat rationality and power asymmetrically. According to these scholars, when experts fail, politics has triumphed. But when experts succeed, credit goes to rationality, not politics. As a result, rationality is believed to be incompatible with interests and displaces power.

The source of this asymmetry is these scholars' implicit acceptance of the "doctrine of objective knowledge", which believes that experts' power is based on the superiority of their intellectual rationality. John Friedmann presented this doctrine elegantly in his article "The Epistemology of Social Practice" using Karl Popper's philosophy of science (Friedmann, 1978).

Friedmann argues that the legitimizing claims of technocracy—or “rule by know-how”—rest with its unique epistemology, which, drawing on the name of a famous work by Popper,

Friedmann calls “objective knowledge”. He described these claims as follows:

Ultimate reality is singular and integral, and its varied manifestations are capable of being known objectively...But the ability to make such assertions differs among people according to their training and vocation. Those whose mastery of the techniques for acquiring objective knowledge is superior, or who have superior access to such knowledge, are also justified in making decisions and committing resources for those whose knowledge is restricted and inferior (Friedmann, 1978: 76).

Friedmann (1978) argues that technocrats “transpose the abstract properties” of knowledge in one “world”—the world of objective thought, which is an autonomous space occupied by thinkers where knowledge production unfolds—to control events in another world—the world of physical states and historical events, which is a managed space occupied by non-professional doers who must obey thinkers’ commands (Friedmann, 1978). This is the model social reformers and critical theorists are drawing on in their respective promotion or denouncement of technocracy. Each group believes in technocracy’s ability to control events in the physical, or “real” world by, for example, constraining politics. Social reformers welcome technocrats’ ability to limit “inefficient” partisan interests and quell dangerous populist passions (Centeno & Silva, 2016; Tugwell, 1954). Critical theorists, on both the left and the right, decry technocrats’ ability to suppress political voice and local knowledge: the left is principally concerned with the threat to cultural and indigenous practices and the right with individual preferences and market prices (Easterly, 2007; Escobar, 2011).

Even within this asymmetrical treatment of rationality and politics, the accounts of expert influence are partial. A principal property of objective knowledge—the scientific method—has been neglected as a source of experts’ power (Hirschman & Berman, 2014). Friedmann (1978) outlines several key properties of objective knowledge. Objective knowledge is universally valid

and ahistorical. The “truth value” of any proposition stakes out a claim that is independent of both time and place. Objective knowledge is expressed in language that is codified and abstract. Access to it is restricted because the special language in which objective knowledge is expressed has to be learned through several years of arduous training narrowly available to a technocratic elite. Objective knowledge is both cumulative and progressive. It is developed and refined on the basis of the rules of scientific method that include standards of truth, content, and validity. This property of intellectual rationality is more akin to praxis than thought (Oakeshott, 1991). This is why Popper said that scientists were like workers who added to the growth of objective knowledge like masons working on a cathedral (Popper, 1972).

Studies of experts are dominated by the first two properties of objective knowledge: universally valid knowledge, or ideas, and codified language, or discourse. By contrast, the third property of intellectual rationality, the practice of scientific method, has received far less attention in studies of state and development experts. The study of the power of expert ideas is in its infancy but growing. Scholars have turned to ideational explanations for political and policy change out of dissatisfaction with dominant approaches in political science such as interest group, rational choice, and historical institutionalist explanations (Mehta, 2013). These scholars focus on the ability of ideas to shape individual interests and organizational objectives (Beland & Cox, 2010). Meanwhile, studies of expert discourse are well-established and discourse analysis is the preferred medium of postmodernist scholars such as postdevelopmentalists and postcolonialists (Crush, 2012; Watts, 1993).

Certain influential philosophers have inspired planning and development scholars to focus on expert discourse. The work of Jacques Derrida stimulated works analyzing experts’ language and critiquing state rationality (Harper & Stein, 1995). Foucault spurred a whole

program on governmentality studies that focused on “discursive regimes” and the power of expert discourse to create truth and shape subjects (Burchell, Gordon, & Miller, 1991). Governmentality studies has been very fruitful for understanding state expertise as a *governmental program*, such as conditional cash transfers (Ruckert, 2009), and power as a set of social relations that draws on state power but can extend far beyond the state’s boundaries (Ilcan & Phillips, 2010; Li, 2007). Studies of governmentality were one of the first and most important contemporary attempts to understand the nexus of knowledge, power, and the work of experts. While these studies help to understand how, for example, expert discourse and techniques can be used to monitor populations and define normalcy and deviance, the program nonetheless falls short of explaining the role of experts and whether their effects had anything to do with knowledge construction inside state organizations. This neglect of state experts and their practices is largely because Foucault had a tendency to dismiss the role of the expert agent and the role of the bureaucracy in power relations (Clegg, 1994; Fraser, 1981; Jessop, 2001).

### **Research Problem and Purpose**

The problem this dissertation addresses is that social reformers and development critics fail to describe particular forms of state rationality as political accomplishments and therefore contingent, contestable, and changeable. Furthermore, they understate methods as a key conduit of expert power. With respect to this study’s focus on methods, it is not a problem that intellectual rationality has been identified as a source of expert power. Rather, the issue is that the focus has been on expert discourse, and to a lesser extent, expert ideas, and those have been treated as *the* sources of expert power. This power is presumed to be based on experts’ identity and relationship with objective knowledge and presented as invariable. As a result, we are left looking at expert power “through a glass darkly” (Moynihan, 2009).

My argument is that one of the central functions of expert methods is to translate broad policy objectives into programmatic and administrative reality and therefore merits close investigation and scrutiny. In order to understand state experts' power we must closely look at some of the work they do most often—analysis for decision-making. We can achieve a fuller and clearer understanding of expert power and its political effects by researching how experts stabilize methods that do work on their behalf.<sup>1</sup> I am not suggesting that methods do not encapsulate ideas and discourses. They do. Governance indicators, for example, cannot exist without new institutional economics and the rhetoric of good governance. Hirschman and Berman (2014) refer to this combination of discourse, ideas, and methods as “cognitive infrastructure”, which includes experts’ “styles of reasoning”—their orienting concepts, ways of thinking, and assumptions—in addition to their methodological approaches. However, existing investigations often stop after explaining how ideas and discourse affect organizations and institutions (e.g., Fukuda-Parr & Hulme, 2011). How ideas and discourse shape experts’ methods, which I refer to as “governing rationales”, and the effects these methods have on organizations should also be examined. This focus on methods can shed light on the “last mile” of expert influence.

With respect to the broader question of whether state experts are powerful, this dissertation specifically asks how certain economic methods, or planning and policy devices (Hirschman and Berman, 2014), get adopted by organizations as governing rationales and in turn affect organizational power relations, decision-making, and the distribution of resources. It is important to study how experts’ methods get established and the effects they have because

---

<sup>1</sup> I use the term *stabilized* throughout this dissertation because *institutionalized* gives a false sense of permanency. I thank Elizabeth Popp Berman for this observation.

orthodox economics, and perhaps other current dominant rationalities, are assumed to be structurally given and inevitable (Swyngedouw, 2014). However, if they are viewed as political as any other project, then they are contingent and contestable. Furthermore, unlike governmentality studies where scholars have limited their focus to expertise's effects on normalization and subject formation, I examine the more pedestrian issues of methods' effects on bureaucratic power relations, decision-making, and the eventual distribution of organizational resources. These issues are equally important but considerably more neglected among Foucauldian-inspired scholars examining the relationship between expertise and power (e.g., Gupta, 1998).

I employ a science studies approach to answer questions of how state experts establish their methods as policy devices and what effects those methods have. I support this broader conceptual approach with modern and postmodern perspectives on power, a diverse genre of sociological literature on quantification processes in organizations, and a practice ontology that focuses on governance strategies that include factors such as actors, techniques, and power rather than on experts' discourse or ideas alone (Best, 2014).

A science studies approach is advantageous for several reasons. First, applying science studies—usually reserved for studying scientists, engineers, or other knowledge workers in laboratories—to state organizations can help us better understand the unique entrepreneurial efforts and political strategies of knowledge-based state experts. Second, science studies treats knowledge production as political work that faces resistance (Latour, 1988). Finally, science studies suggests that methods “black-box” social science ideas and discourse (MacKenzie, 2005).

To study experts' methodological projects or particular forms of state instrumental rationality, I employ a value-rational or phronetic research approach. A value rational approach allows me to keep the focus on experts' power and "micro-practices", such as methods (Flyvbjerg, 2004). I ground the investigation of how state experts stabilize their methods and the effects those methods have within an embedded case study of a United States (US) global development finance agency and use a mix of qualitative and quantitative evidence drawn from interviews, documents, and datasets. The embedded case study concentrates on three economic methods used to make decisions in the US Millennium Challenge Account (MCA), a new global development finance program, and the agency tasked with its implementation, the Millennium Challenge Corporation (MCC). These methods—governance indicators, growth diagnostics, and randomized controlled trials (RCT)—are the dominant techniques used for allocating development finance, designing country development programs, and evaluating development projects, respectively.

### **Research Design and Context**

**Research Design.** Jacqueline Best's (2014) excellent overview of contemporary global development finance describes the major strategies that development finance agencies use in the field. She describes four "governing strategies"—country ownership, governance standards, risk mitigation, and results measurement—that have come to define the field in the so-called Post-Washington Consensus era. Whereas Best (2014) analyzes these strategies and practices across the breadth of the field by looking at multiple institutions such as the International Monetary Fund, multilateral development banks (MDB), and bilateral aid agencies, my dissertation focuses on a specific organization and looks in-depth at specific methods that support some of these governing strategies. I chose an embedded case-study research design for its ability to isolate, as



reasonably as possible, economic methods' effects; permit enough contextual detail to allow naturalistic generalization; and enable a mixed methods research approach. Moreover, while both Best (2014) and this dissertation focus on a very important aspect of global development—development finance policymaking—I examine this “meso” level process through the lens of governance indicators and also look at more “micro” level development processes such as national development planning (growth diagnostics) and local project evaluation (RCTs). This vertical embedded case study design examines the development process from start to finish through the lens of a single agency and is rare among studies of development expertise.

Both the case of the MCA-MCC and these three development economic methods were selected because of their paradigmatic qualities in the global development finance field. Governance indicators represent one way for development agencies to operationalize their focus on client countries' institutions to increase “aid effectiveness”. Growth diagnostics help demonstrate these development agencies' commitment to country ownership, which acknowledges countries' unique economic, political, and social circumstances during development planning. RCTs can contribute to institutions' results agendas that aim to put the field on sound methodological footing and demonstrate success to its critics.

**Research Context.** The Millennium Challenge Account is a dedicated development fund as part of the United States International Affairs budget under the category of Bilateral Economic Assistance. The MCC is both a bilateral aid and global development finance agency. It occupies a space between fields: the field of US foreign assistance and its related organizations such as the United States Agency for International Development (USAID) and the field of global development finance and its related organizations such as the World Bank. The MCA-MCC was

chosen as a case study because it represents a paradigmatic case of changes in global development finance governance during a period of momentous change.

The MCA was announced in March 2002 in Monterrey, Mexico at the first-ever United Nations Conference on Financing for Development. Over fifty Heads of State and two hundred Ministers of Finance, Foreign Affairs, Development and Trade attended. They produced the Monterrey Consensus, a resolution stating that developing countries should take a greater leadership role in their development through good governance and the mobilization of domestic resources and developed nations should substantially increase resources for global development assistance (Annan, 2002). These appeals by the global community, and lower-income countries in particular, pushed the United States to increase its official development assistance. Another reason for this increased commitment included the George W. Bush Administration's belief that addressing global poverty—thought to be a catalyst for terrorism—would reduce the chances of attacks such as 9/11 as part of a broader National Security Strategy (Bush, 2002b).

The MCA was originally planned for US \$5 billion per year starting in 2006. Actual appropriations amounted to far less; Congress has appropriated only one-fifth of the original request, or US \$1.2 billion on average, for every year between 2004 and 2014 (Tarnoff, 2015). Globally, the increase represents approximately the same size as the US annual contribution to the Asian Development Bank's Asian Development Fund and Switzerland's annual bilateral development aid budget. Nevertheless, the MCA represents one of the most significant increases in US bilateral development assistance since the John F. Kennedy Administration (1961–1963) established the Alliance for Progress in 1961 (Radelet, 2003). President Bush signed the *Millennium Challenge Act* into law in January 2004 as part of the *Consolidated Appropriations*

*Act of 2004* after bipartisan passage in the US Congress. The Act created the Millennium Challenge Corporation, a federal government corporation, to execute the MCA.

The MCC was designed to be an independent government agency. It has a Chief Executive Officer (CEO) but is governed by a nine-member Board of Directors rather than a single departmental head such as the Secretary of Labor or Energy. Of MCC's nine-member Board of Directors, four members are from the non-profit or private sector. The agency is also exempt from select federal statutes, particularly in relation to the hiring and status of employees and procurement regulations, which is common among federal government corporations (Government Accountability Office, 1995). The label "independent" is a misnomer, however. In the most important respects, the MCC is a public US government entity. Most federal government corporations are constitutionally dependent on the President and MCC is no different (Breger & Edles, 2000). The President appoints the majority of the Board of Directors including MCC's CEO and controls its budget request to Congress. In turn, MCC is wholly dependent on Congress to appropriate some or all of that request every fiscal year. Thus, the MCC is similar to federal government corporations that came before it such as the Tennessee Valley Authority and Overseas Private Investment Corporation (Kosar, 2011).

The MCC was intended to be a beachhead organization to move the field of US foreign assistance closer to the image of multilateral development banks in development finance. Many Bush Administration officials involved in its creation spoke about an "MCC model" and the need for a demonstration effect. They also wished to establish a new agency that would represent an incubator for ideas and strategies that they wanted to see proliferate in development finance as the field transitioned from the "Washington Consensus" paradigm to the "Post-Washington Consensus" era. Some of the highest-ranking officials in the Administration's National Security

Council, Council for Economic Advisors, and Cabinet such as Condoleezza Rice, Glenn Hubbard, and John Taylor were directly involved in spearheading both the MCA-MCC and global development finance reform (White House, 2002; Taylor, 2002b).

The “Washington Consensus” is a term that symbolizes the structural adjustment policy paradigm of the influential international financial institutions’ (IFI) in Washington, DC such as the International Monetary Fund, US Treasury, World Bank, and Inter-American Bank during the 1980s and 1990s (Babb, 2012). The Post-Washington consensus, not a consensus in the earlier sense of the term, includes a set of ideas and strategies preoccupied with freeing nations and individuals from “poverty traps”. Most mainstream development economists believe that the Washington Consensus’ private property-based, market-oriented reforms were necessary; they just needed to be supplemented with prior interventions at the level of nations and individuals (Reddy, 2012). The task of the macroeconomic diagnostician was to identify what is required for a country to overcome its poverty trap (Rodrik, 2009; Sachs, 2006). Microeconomists must find “what works” at the individual and community level so that the poor can take advantage of market opportunities. This “Washington Consensus plus” position critiques the early paradigm for its insufficiency and “focus[es] on what is necessary to alleviate obstacles to spontaneous and self-sustaining growth and development based on the grasping of market opportunity” (Reddy, 2012: 4).

At this same time, the discipline of development economics was undergoing transformation. Throughout the 1980s and into the 1990s, development economics, with its statist approaches to modernization based on models of stages of growth and surplus labor, struggled to discover a new identity after being pushed aside in American economics departments by free market macroeconomic orthodoxy (Hirschman, 1981). At the turn of the

century, new empirical approaches helped development economists experience a comeback. Their reliance on more “relevant” growth models for developing countries and preference for empiricism over formalism, both of which earlier marginalized them in academia (Bøås & McNeill, 2004), were seen as strengths in a post-paradigmatic Post-Washington Consensus period. A number of new methods, principally the three studied here, emerged to tackle key questions that were also amenable to how governance strategies were changing in global development finance. The MCC existed at the crossroads of these changes. In other words, it represented the new face of development governance and applied development economics in the Post-Washington Consensus era.

### **Conceptual Framework**

The conceptual framework for this dissertation is based on applying three science studies concepts—the principle of symmetry, trials of strength, and opening the black box—to the study of state experts and their methods. These concepts are buttressed by theories of power and sociological approaches to quantification.

Sociologists of science who claimed that historians approached the investigations of “successes” and “failures” of science asymmetrically advocated for the principle of symmetry. The principle of symmetry suggests that both successes and failures should be treated equally, approached with sociological explanatory principles and methods, and have their sociological, rational, irrational—and I argue political—elements. For example, Yonay (1998) did not accept the textbook version of the history of economics where the contemporary postwar neoclassical school is a natural outgrowth of the prewar classical school based on the superiority of its mathematical modeling of marginal costs and prices. The institutional school was equally compelling but lost to the neoclassical school as the two schools’ groups engaged in a social and

political “struggle over the soul of economics” (Yonay, 1998). The principle of symmetry dispels technocrats’ narrative that they are the bearers of rationality able to resolve conflict because their special relationship with truth favors improving collective welfare rather than the interests of various constituents (Centeno, 2010). Furthermore, it investigates the political work of economic experts, which shows that the “post-political” neoliberal condition, which is closely associated with orthodox economic rationality, is not a given or inevitable state.

I combine with the principle of symmetry theories of power to explain experts’ politics. Framing politics as power distances us from the depoliticization and post-political rhetoric that frames politics narrowly as resistance to the state. Such literatures position state domination on one side and strategies for counteraction on the other (Scott, 2008). However, if politics is understood as power, and as Foucault suggests, all social relations are relations of power (Foucault, 2001), then experts are implicated in power relations as well. They are not simply the bearers of rationality—a property they either do or do not possess based on their training, credentials, and reputation—but are complex agents able to exercise various strategies of power in their relationships.

I approach the state as an arena of competing experts with their respective rationalities (i.e., paradigms rooted in episteme or praxis) by applying science studies’ concept of trials of strength. Experts, like scientists, have to conduct their work and defend it from critics, particularly other state expert groups with conflicting rationalities. Paradoxically, even while development is a deeply contested topic in academia, scholars researching development expertise in the state often treat it monolithically. In doing so, they also inadvertently cast the state as having a unitary logic (e.g., Goulet, 1986). This is a product of neo-statists’ attempts to dispel views of the state as “society-centric” (the state as one among many outcomes of structural social

forces) with arguments for state autonomy and Weberian theories of bureaucracy that separate state and bureaucratic rationality from other rationalities such as the practical rationality of politics or local communities (Jessop, 2001).

Methods become black-boxed when debates over the procedures of (social) scientific practice appear settled and their internal workings are overlooked. Opening the black box of methods can reveal experts' intermediate choices and political values, which, when methods are stabilized as policy devices, can engender political effects (MacKenzie, 2005). A genre of sociological literatures that falls under the broad rubric of "sociology of quantification" helps us "get inside" a wide array of expert methods to understand their biases and effects (e.g., Espeland & Stevens, 2008). I ask how expert methods make decisions and how those decisions affect organizational power relations and the distribution of resources.

### **Significance of the Research**

**Research Significance.** This dissertation makes a contribution to the debate over the power and influence of experts in planning and development. It fills a gap in the literature by examining experts' efforts to establish their methods and the power those methods have. As mentioned earlier, the prevailing belief about expert power in planning and development is that intellectual rationality is experts' primary source of power, this power is strong and stable, and the power of intellectual rationality is drawn from its ideas and discourse. Without political power, however, experts cannot realize the power of their intellectual rationality and without looking at methods we cannot understand the full extent of the power of intellectual rationality

This dissertation is not the first to question this prevailing belief. While the presumed power of intellectual rationality represents the dominant framework, other theoretical and empirical works both within and outside of planning have revised this position. Some studies

have suggested that experts are not as powerful as many think. Elite theorists in the social sciences have argued that experts are not powerful in themselves but are rather instrumentalized by political, economic, and interest-group elites (Mills, 2000; Reed, 2012). Some heterodox planning theorists agree that experts' power is exaggerated, but not for the reasons elite theorists put forward; experts' intellectual rationality is no match for rationalizations made by actually powerful actors (Flyvbjerg, 1998). Others, by contrast, believe in the power of intellectual rationality but only under conditions of "expert autonomy"—experts must liberate themselves from bureaucratic and legislative authority in order for their rationality to prevail (Bimber, 1996; Dargent, 2011).

This dissertation argues that experts are powerful because they establish methods that do work on their behalf (Eyal & Levy, 2013). In other words, experts' intellectual rationality is powerful, but not in the ways we originally thought. They must first establish their methods as policy devices, which requires them to exercise power thought to be the reserve of politicians such as legislative representatives or bureaucratic officials. Thus, methods, as a form of intellectual rationality informed by ideas and discourse, are a strong but not necessarily stable form of power. Experts must do political work to stabilize methods and this makes their methods contingent and contestable.

**Political Significance.** The effects of the asymmetrical treatment of experts and politics extend beyond a proper scholarly accounting of state expertise; it frames political action as well. As it stands, there is a problematic contradiction within liberals' "rational dream" of advancing human welfare through a state guided by science, reason, and expertise (Toulmin, 2009). Achievement of the rational dream undermines two other deeply held liberal values: equality and neutrality (Turner, 2001, 2003). State expertise threatens equality of voice in a



liberal democratic polity that strives to afford equal weight to all views. It also threatens the liberal state's goal to remain neutral in the face of competing viewpoints. This engenders a technocracy-democracy divide where technocracy and democracy are locked in a zero-sum struggle. As Collins and Evans (2009) put it: "Democracy cannot dominate every domain, that would destroy expertise, and expertise cannot dominate every domain, that would destroy democracy" (Collins & Evans, 2009: 8). Consequently, for postmodernist critics of experts the principal response has been to decentralize and diversify knowledge through popular epistemology. Postmodernists reject the notion that the state or science has some privileged position in society or relationship with truth; everyone's knowledge is equally important and valid and deserves to play a role in planning and governance (Fuchs, 1997; Harper & Stein, 1995).

The postmodernist reaction to state expertise is theoretically and rhetorically appealing but when taken to its logical conclusion results in a "postmodern abyss"—an undermining of the intellectual base of the modernist planning project without an alternative guide for the role of knowledge for public action (Beauregard, 1991). Postmodernists risk being handmaidens to regressive views by discrediting state expertise without presenting alternatives (DuPuis & Gareau, 2008). As multiple philosophers of expertise have pointed out, state experts are not going anywhere for the foreseeable future because they contribute to essential decision-making tasks such as complexity reduction and knowledge aggregation (Selinger & Crease, 2006; Turner, 2013). Turner (2008) rhetorically asks, "Should development economists just shut up"? Turner answered this question by saying:

Realistically, they don't know how to produce vibrant economies in the Third World or lead millions out of poverty under the actual political and cultural conditions of impoverished nations. Does this mean they should stop trying as experts to formulate policies and policy ideas? Probably not... This is perhaps a domain in which the urgency

is such that trying on the basis of very limited understanding is the only option (Turner, 2008: 40).

Examining how certain experts, like orthodox economists, have gained power and dominant methods, like orthodox economic rationality, come to be stabilized and have specific effects is meaningful for planning and political practice in two ways. First, it encourages reflective practice that can lead to subsequent reform if so desired. If we discover that methods influence the ends of planning and administration, we can debate whether the effects we see are those that we want. Technocrats are complex actors with experience and values and many are open to reform (Roy, 2012; Wilson, 2006). Accordingly, rather than addressing the “democratic-deficit” resulting from technocratic governance through popular epistemology alone, we can pursue democracy by “getting inside” the techniques (Latour, 1998). Second, and more importantly, one modest route out of the post-modern abyss is to investigate the state as a plural arena of competing rationalities and interests. Describing how dominant experts win trials of strength demonstrates that the state is an arena of competing rationalities where other experts can likewise establish their methods. Orthodox economic governing rationales, and perhaps other currently dominant approaches, are not structurally given and inevitable. They are political projects like any other and thus contingent and vulnerable.

### **Overview of the Dissertation**

This dissertation is organized according to the following chapters: Literature Review (2); Research Design and Methodology (3); one chapter each on case studies of Governance Indicators (4), Growth Diagnostics (5), and Randomized Controlled Trials (6); Interpretation of Findings (7); and Conclusion (8). The Literature Review explains the conceptual framework and significance of the research by surveying the major debates surrounding three sources of expert power—their intellectual rationality, institutional relationships, and methods. The views on

these sources of expert power are drawn from diverse literatures such as planning, public policy, and political science. The instances where the field of science studies has implicitly addressed shortcomings in other dominant genres are identified. A presentation of the major theories of power from both modern and postmodern perspectives supports these genres.

The Research Design and Methodology chapter (3) describes the dissertation's research approach, design, methodology, data sources, and procedures used to answer questions about the power of economic expertise. The dissertation questions are approached through an embedded case study design of the economic methods involved in creation of the MCA and how the MCC designs and evaluates development investments. The embedded research design was chosen for its ability to allow for a mixed methods research approach. The data sources include both primary data sources in the form of 50 interviews and secondary data sources in the form of hundreds of documents and multiple quantitative data sets.

The Governance Indicators case chapter (4) describes the rise of governance indicators and the Bush Administration's adoption of a particular set of indicators, the World Bank's Worldwide Governance Indicators (WGI), for its aid effectiveness strategy. It then describes how this rationale of using indicators for selecting countries, promoted by the economic policy community, came into conflict with the foreign policy community's geopolitical rationale and what state economic experts did to overcome the foreign policy community's opposition. Finally, the chapter presents the MCA's country selection criteria and the WGI's effects on how low-income countries' institutions are depicted in development finance decisions and the distribution of MCA resources. The Growth Diagnostics case chapter (5) describes the origins of growth diagnostics and how they transformed the development-planning environment at the MCC. It then discusses how economists went about building a diagnosis and how the growth

diagnostic rationale conflicted with social, gender, and environmental experts' rationale for how country development programs should be designed. The chapter closes with an explanation of how growth diagnostics reshaped the participatory planning process and the composition of development programs. The RCT chapter (6) describes the rise of RCTs, the opposition MCC evaluation economists faced when introducing RCTs to MCC, and how RCTs were disproportionately applied to agricultural and rural development projects. It explains how evaluation economists overcame opposition to randomly assigning participants to projects, which was irreconcilable with implementers' desire to target participants purposely according to sociological theories of diffusion. The chapter concludes by describing the effects of RCTs on the composition of project participants and project implementation.

The Interpretation Chapter (7) describes the main findings from the three empirical cases. The cases combine empirical evidence of how economic experts achieved autonomy and overcame opposition from other expert groups with modern and postmodern theories of power. I find that economic experts' actions were motivated as much by political interests as any other group because they had an interest in stabilizing their methods as governing rationales. Also, economists' political work required the exercise of power because the more opposition the economists' method faced the more power they had to exercise to stabilize their methods. Finally, when both politicians and experts transferred their decision-making authority to methods, and those methods became codified in legal agreements, methods had political effects that were not simply discursive. These effects included the reshaping of bureaucratic power relations, a shift in decision-making processes and actors, and the distribution of global development finance resources that point to a need to combine modern and postmodern concepts of power rather than rely on a single approach alone.

The Final Chapter (8) describes the implications of the study's findings for planning research, practice, and development policy. I contend that a science studies approach can benefit other planning scholars studying state experts and their methods. The chapter also describes what the three development economic methods studied here mean for development policy with respect to participatory planning and inequality. It goes on to explain in detail how revealing methods' political values and assumptions can instigate reflective practice. Importantly, the final chapter describes what the reframing of experts and methods in this study could mean for practicing planners, particularly advocacy planners, as the discipline continues to grapple with the postmodern abyss. The chapter closes with a discussion of the limitations of this study and suggestions for future research.

## Chapter 2: Major Debates over the Power of Experts

### **Introduction**

Scholars who have studied experts have debated their power and its basis in their knowledge or intellectual rationality. Scholars have also examined the relationships between experts and legislative and bureaucratic authorities and, moreover, asked whether “expert autonomy” was essential to expert power and, if so, how experts attained it. Finally, scholars have debated the importance of experts’ governance methods, also known as policy instruments or devices; as this chapter shows, some having dismissed the weight of expert influence over methods while others more recently have suggested that the power of expert methods has been overlooked and deserves closer attention. These scholars have not engaged in direct debate over experts’ sources of power, but their juxtaposed work highlights their contrasting views. Scholars of science studies have often implicitly addressed shortcomings in other dominant genres. This chapter highlights three instances when this has occurred.

To narrow the review’s scope, I focus principally on experts in or directly associated with the state rather than look at professions or lay expertise. Finally, most of the examples and empirical references are drawn primarily from studies of economic expertise and supported by examples from development studies.

### **The Asymmetrical Treatment of State Experts**

**The Rational Dream and Its Critics.** A group of scholars in planning theory, the policy sciences, and public administration known broadly as “social reformers” are driven by faith in and commitment to the rational dream (Boyer, 1986; Friedmann, 1987; Hall, 2002; Tugwell, 1954). This rational dream is an enduring conviction, which dates back to the ideas of Francis Bacon and Saint Simon, that scientific, rational planning has the power to overcome politics

(Baum, 1980; Turner, 2003). This linear model is predicated on a division between a world of values, occupied by decision-makers who grapple with problems, and a world of facts, occupied by experts who analyze problems and propose solutions based on their specialized knowledge and skills (Baum, 1980; Weiss, 1979). Within this division, the authority of experts is intimately tied to a body of abstract theoretical knowledge that sociologist and theorist Max Weber referred to as intellectual rationality (Altshuler, 1965a; Weber, 2009). Intellectual rationality is a source of experts' claims to legitimacy and governing authority. This form of authority differs from administrative authority claimed by executives, constitutional authority claimed by political officials, and ideal-based authority typical of interest groups (Brint, 1990).

Social reformers implicitly assume that experts are uniquely suited to “end ideology” because of the way the linear rational model frames experts' relationship with knowledge and the state (Bell, 1965; Friedmann, 1987). Experts are powerful because they can blunt the parochialism of partisan interests and populist passions by speaking truth to power (Mannheim, 1950; Wildavsky, 1979). In other words, they mobilize “truth”, or scientific knowledge, *to* power, that is, to the real decision-makers such as administrators and elected officials. Social scientists are presented as the bearers of rationality, not powerful agents, decision makers, or actors with personal interests. Here, knowledge is largely understood as information that represents the public interest or that is delivered at strategic moments as advice intended to inform decision-making and policy processes (Altshuler, 1965b; Baum, 1980). When this model breaks down, the power of expertise evaporates.

Neo-marxist geographers, postdevelopmentalists, and other critical theorists oppose the rational dream but nonetheless implicitly share social reformers' normative and theoretical positions. These scholars believe that the authority of experts derives from their relationship

with and access to knowledge and the state (Escobar, 2011; Sachs, 1992). They believe, reflecting Max Weber's theories of bureaucracy, that the state is the locus of technical rationality, which can be mobilized and put to use in a brutally efficient way (Mitchell, 2002; Scott, 1998). Goal-oriented professional technocrats, including social scientists whose legitimacy derives from the epistemological foundations of modern science, administer this technical rationality (Breslau, 1997; Eisenstadt, 1958). For example, in his discussion of contemporary environmental politics, Swyngedouw (2009) states that technocratic governance "is a politics legitimated by a scientific consensus which, in turn, translates into a political consensus" (2009: 602). According to Slavoj Žižek (2014), it is a politics that "legitimizes itself by means of a direct reference to the scientific status of its knowledge" (2014: 211).

Like social reformers, critical theorists believe that expert-led governance constrains politics. One group of neo-marxist critical theorists believe that the post-Cold War neoliberal political economy and the performance-driven managerial state constitute a structurally predetermined state of affairs (Beveridge, 2017; Swyngedouw, 2014). To them, this technocrat-led administrative state operating in the name of free markets is depoliticizing (Beveridge & Koch, 2017). This is because technocratic governance forecloses the space in which alternatives to the so-called neoliberal consensus can emerge. Inside the institutions responsible for neoliberalism, reformers either become co-opted or legitimize the consensus (Beveridge, 2017; Swyngedouw, 2005). In other words, past the point of viable political alternatives, we become *post*-political. Consequently, the only legitimate political activity that remains is protest politics that exists outside of, and acts antagonistically towards, consensus-driven actors and institutions (Beveridge & Koch, 2017; Flinders & Wood, 2014).



Postdevelopmentalist scholars think similarly to post-political theorists. Development experts cast inherently political questions and issues, such as access to land or public services, in technical terms (Ferguson, 1994). Thus, they regard the expansion of the development state as a depoliticizing force, one that creates a situation of *anti*-politics (Ferguson, 1994; Sachs, 1992). Here, real politics takes place beyond the boundaries of the state and its experts (Esteva & Prakash, 1998). Postdevelopmentalists equate politics with political challenges to the development state (Kiely, 1999). These scholars' stance reinforces the enduring divide between rationality and politics in political and planning theory. Postdevelopmentalist theories frame politics narrowly and conceive of relationships between state experts and the public in binary terms. Thus, the only way to politicize governance and development is to engage in participatory or protest politics (Turner, 2003). When such activities are successful, state experts are (partially) dethroned, and governance and development become political again.

**The Chastened Dream and Limits of Expertise.** While the social reformers' enduring faith in "knowledge as power" and the leftist critical theorists' narratives of expert-led depoliticization give us the impression that experts are influential, a third group claims that rationality is in fact a weak form of power and no match for "realrationalitat," or the rationalizations of actually powerful agents (Flyvbjerg, 1996). From this perspective, expert advice and knowledge is routinely ignored or dismissed by those with *real* power. Studies of expert-state relations and knowledge utilization at the municipal level concluded that expert knowledge has little or no effect on decisions and that experts are inconsequential (Collingridge & Reeve, 1986). A case study of sustainable urban development in Aalborg, Denmark found that rationality is a weak form of power and is no match for rationalizations made by the

powerful (Flyvbjerg, 1998). If rational methods are to be effective, they have to conform to existing political paradigms and agendas (Innes, 1990).

Political scientists and sociologists who espouse an elite theory of the state believe that powerful elites instrumentalize experts, who become mere “servants of power” (Brint, 1990). That is, experts become “window dressing” to justify decisions made by politicians for non-technical reasons (Banfield, 1961). For example, John Kenneth Galbraith argued that the technocratic elite would either prove unable to challenge political leadership or be transformed by its participation in governance (Galbraith, 1972). There is evidence to support the elite theorists’ view. Indian planners in the immediate aftermath of independence suggested that foreign economic advisers provided “airforce cover for the army”—not suggesting a way forward so much as providing reinforcement for what had been already planned (Engerman, 2013). Tandler (1977) cited the use of economists’ cost-benefit analyses to support foreign-aid investment decisions after the fact. Experts and planners who decry their lack of voice and influence in decision-making and governance add to elite theorists’ views (Baum, 1980; Lindblom, 1979). These positions and analyses give the impression that experts and their rationality are failures and that power has prevailed. This impression has led to what Jal Mehta (2009) calls the “chastened dream”.

**The Asymmetrical Treatment of Experts.** The literatures described thus far make it appear as though experts have no politics of their own. Not only does this occlude the myriad ways that experts actually inform the policy process, it allows the continued depoliticization of experts and their actions. Both the reliance on instrumental-rational knowledge and the desire to keep knowledge and politics separate—what Latour (2012) calls the “modernist settlement”—has led to idealized conceptions of the roles of experts and politicians and how expert knowledge

should be applied to policy problems (Baum, 1980; Grindle, 1977). Furthermore, when the presence of experts increases in planning and administration, criticism arises that development has become depoliticized or dominated by technical approaches (Wilson, 2006). As a result, experts are treated asymmetrically. When the linear rational model breaks down, social reformers conclude that politics, or real power, has won. When the linear model is most ardently followed and experts are in charge of planning and administration, critical theorists decry that planning and administration have become depoliticized. Either way, experts' work is not cast as political.

**The Principle of Symmetry in Science Studies.** Asymmetrical explanation has long been a problem in so-called Whig histories of science. Sociologists of science claimed that Whig historians and others approached true beliefs, or science's "successes", with rationalist explanations and approached false beliefs, or science's "failures", with sociological explanations (Sismondo, 2011). To correct this tendency, several sociologists of knowledge at the University of Edinburgh advocated the principle of symmetry (Bloor, 1991; MacKenzie, 1981). The principle of symmetry suggests that science's successes, or its beliefs thought to be true, should be explained with factors such as ideology, interests, and politics, which are usually reserved for science's failures and beliefs thought to be false. In other words, the same factors are involved in the production of truth as in the production of falsity (Bloor, 1981). The principle of symmetry is most closely associated with sociologist David Bloor, who outlined "four tenets" for the sociology of scientific knowledge. Two of the most important points stipulate that the sociology of scientific knowledge should be: 1) impartial with respect to truth and falsity, rationality and irrationality, and success or failure; and 2) symmetrical in its style of explanation, so that the same types of causes would explain both true and false beliefs (Bloor, 1991).

A paradigmatic case of the application of the principle of symmetry is *Leviathan and the Air Pump*, where the success of chemist Robert Boyle's experimental methods was understood as a social and political achievement over the powerful philosopher Thomas Hobbes' prevailing plenist ontology (Shapin & Schaffer, 2011). Shapin and Schaffer (2011) treat truth, objectivity, and scientific adequacy as accomplishments, and the problem of generating and protecting knowledge as a political problem. Likewise, they engage in symmetrical explanation by taking a "stranger's" or outsider's perspective of Boyle's characterization of natural philosophy or Hobbes (simply a "failed experimentalist"). Shapin and Schaffer (2011) stated: "Our goal is to break down the aura of self-evidence surrounding the experimental way of producing knowledge, and 'charitable interpretation' of the opposition to experimentalism is a valuable means of accomplishing this" (Shapin & Schaffer, 2011: 13). Meanwhile, in a paradigmatic case study of rationality and power in planning in Aalborg, Denmark, Flyvbjerg (1998) describes the city bureaucracy's attempts to implement an award-winning plan and the Aalborg Chamber of Commerce's successful attempts to alter the plan to suit its interests. In Flyvbjerg's depiction, the social reason among Aalborg planners represented rationality and the Chamber's opposition—or private reason—represented irrationality (Turner, 2008). Flyvbjerg characterized the Chamber of Commerce as exercising power but did not consider the bureaucrats' plans as expressions of power. Thus, Flyvbjerg was engaging in the type of asymmetrical explanation that is common in planning scholarship (Turner, 2008). The principle of symmetry reconceptualizes the role of experts from one of speaking truth *to* power to one of constructing rationality *as* power.

## **Experts and Theories of Power**

If rationality is politics by other means and politics is a code word for power (Baum, 1983a) then expert rationality is arguably a form of power. I now review what the key theories and literatures on power say about experts. The literature on power is extensive but its approaches can be roughly divided into two camps: modern and postmodern. These approaches are often referred to respectively as “power over” and “power to” (Clegg, Courpasson, & Phillips, 2006).

**Modern Approaches to Expert Power.** Three concepts are central to modern approaches to power: hierarchy, authority, and constraint. In modern approaches, the distribution of power is known—it is concentrated at the top. For Thomas Hobbes, one of the earliest theorists of power, the first line of authority was the Leviathan (Clegg, 1989). The organizing power of the monarchical system was the individual’s relationship with the sovereign, rather than, say, with other individuals. Max Weber, one of the earliest and most important theorists of the state and bureaucracies, posited that hierarchy was instrumental to understanding power, which was concentrated at the top of the state’s vertically organized institutions (Weber, 2009). Weber defined power as the probability of an actor carrying out his or her will even when sh/e is opposed (Uphoff, 1989). Authority is central to this idea of power and has its basis in coercion. Authority is dependent on a special position or role that enables the person occupying it to make commands. This is usually justified by the right to act based on the property of an office. The Romans called this *potestas* (Höpfl, 1999). Even in the case of *potestas*, the office holder must be seen as legitimate, or have the capacity to inspire respect. The Romans called this quality *auctoritas*. Those without *auctoritas* are not “right and proper” and the objects of an officeholder’s command may not see an obligation to obey (Höpfl, 1999). Those with authority

can exercise constraint. Constraint is the ability to determine the alternatives open to others either by direct force, which involves the use of negative physical sanctions, or inducement, which involves sanctions and rewards to influence the calculations and decisions of others (Scott, 2008). Economic resources are central to modern power. Actors who can wield resources to induce groups through sanctions and rewards are powerful, even if they do not possess the legitimate authority of an office. Likewise, authority is a matter of degree and is greater when combined with economic resources (Uphoff, 1989).

Modern theorists believe that monopolization of information and specialized forms of knowledge afford experts power through resources and authority. By gaining and controlling access to information and knowledge, they can equalize power relations by wielding their resources against others (Benveniste, 1972; Turner, 2013). This gives experts the power to sanction or reward. For example knowledge monopolization is achieved through licensing and credentials (Freidson, 1988; Larson, 1979). Scholars in the sociology of professions study the process through which experts professionalize knowledge, making it a scarce resource, as a means to gain authority (Eyal, 2013). This authority requires a demarcation between professional and lay knowledge. How this boundary is created matters less than how the boundary improves the claims of professionals to authority and resources (Gieryn, 1983). Through this process, experts gain legitimacy and, among their audiences and clients, create the sense that their knowledge and counsel is valid, reliable, and should be followed (Scott, 2008). One limitation of the sociology of professions is that it does not address the issue of social relationships among experts either inside or outside of the state.

Another possibility is that experts can become elites (Zald & Lounsbury, 2010). Some experts can reach the “command posts” of hierarchical organizations and bureaucracies such as

central banks, ministries of finance, environmental protection agencies, public utilities, and non-state organizations with a technical focus. For example, economists have recently risen to high-level positions across various branches in Latin American states (Markoff & Montecinos, 1993). These experts are afforded both control over knowledge and professional legitimacy as well as the power to issue legitimate commands because of the office they occupy (Scott, 2008). They are in the minority, however. Most experts work as mid-level bureaucrats at the command of their hierarchical superiors (Coats, 2001).

**Postmodern Approaches to Expert Power.** Postmodern power is concerned more with networks than hierarchy, strategy than authority, and practices than resources. Postmodern power is a non-hierarchical form of power; there is no single entity “pulling the strings.” One actor affects another even if s/he is not some authority with an endowed form of power (Burchell, Gordon, & Miller, 1991; Clegg, 1994). The philosopher Michel Foucault, a major influence on postmodern approaches to power, was concerned with power in “institutionalized strategic complexes,” such as hospitals, prisons, schools, and factories, and their everyday relations with people (Heiskala, 2001). For example, power operates through a network of connected actors, such as attorneys, judges, wardens, and guards in a penal system, rather than commands coming down from a warden’s office (Foucault, 2012).

The roots of postmodern power lie in the thought of Niccolo Machiavelli rather than that of Thomas Hobbes. To Machiavelli, power is derived from strategy, not coercion. Deals, negotiation, fraud, and conflict play central roles in securing order by a strategically-minded prince (Gilbert, 1984). Actors’ strategies are often aimed at building coalitions of allies. Postmodern theories of power are concerned with what power does, not what power is (Clegg, 1989). This power operates through micro-social practices rather than economic resources

(Foucault, 1978). A principal practice in Foucault's work is discipline. Disciplinary power through practices, such as "the gaze", is productive rather than coercive: postmodern power is ultimately concerned with the production of subjects (Foucault, 2001, 2012). Such power can create, for example, docile bodies. Ultimately, the targets of postmodern power willingly conform to power without coercion (Burchell et al., 1991).

With respect to knowledge, postmodern theories believe experts gain power through generosity or the sharing of their knowledge and practices rather than monopoly.

Monopolization of knowledge is isolating and isolated experts are not powerful. The goal of experts is not to restrict access or achieve legitimacy but to graft their work onto what others are doing (Eyal, 2013). This work involves the dissemination of ideas and discourses. Ideas claim to describe the world, the causal relationships, or the legitimacy of certain actions (Parsons, 2002). For example, one popular idea is that government spending will stimulate economic growth (Béland, 2010). The power of experts' ideas lies in their ability to construct both general and particular social and political interests. They can help convince actors that the existing state of affairs is inherently flawed and that major changes are necessary in order to solve its problems (Beland & Cox, 2010). Several case studies in development have shown how economists' ideas have changed the missions of the World Bank, Economic Commission for Latin America, and United Nations Development Program at particular moments in time (Cooper & Packard, 1998; Fukuda-Parr, 2011; Sikkink, 2012).

Discourse involves linguistic and symbolic practices such as narrative, rhetoric, and dialogue, which are expressed through speech, text, visual representations, and cultural artifacts that create new phenomena (Wetherell, Taylor, & Yates, 2001). For Foucault, discourse is more than language; it is a system of representation that governs the way a topic can be reasonably



talked and reasoned about (Mills, 2012). This “discursive regime”, when coupled with power (the couplet “power/knowledge”), generates truths (Foucault, 1980). In order for something to be established as fact or as true, other equally valid statements have to be rendered invalid, denied, or categorized as taboo. Together with statistics (i.e., organized numerical data), discourse constructs the “normal” (Mills, 2012). By defining what is normal, experts can determine and govern what is wrong or deviant (Foucault, 1978). For example, expert discourses decide who is undeveloped or poor by creating norms of (western) progress (Escobar, 2011; Watts, 1993). Experts’ statistics then inform where and how development programs should intervene (Li, 2007). Their use of discourse and discipline is a major focus of governmentality studies, which emerged in the 1980s as an important early effort to fuse knowledge, power, and the work of experts (Burchell et al., 1991).

According to postmodern theorists, experts do not need to become elites to be influential. These theorists focus more on “distributed agency”—that is, the “multiplicity” of intervening agents in a network. Some of the most important networks form epistemic communities (Eyal & Buchholz, 2010). An epistemic community is a network of experts or professionals with recognized competence in a particular domain and an authoritative claim to knowledge within that domain or issue area (Haas, 1992). It is an “invisible college,” whose members span academia, different arms of government, non-governmental organizations, policy circles, and think tanks. Their members work together collaboratively and share principled beliefs and values (Haas, 1992; Maxwell & Stone, 2004). Ultimately, however, the social character or individual positions of their members is less important than the ability of its members to be effective channels for the circulation of new ideas from the communities to governments (Eyal & Buchholz, 2010).

**Expert Autonomy and Expert Power.** If postmodern scholars of power focused on state experts and their generosity of knowledge and membership in knowledge-based networks, others who conceive of power in more modern, hierarchical terms sought to comprehend the autonomy of state experts from legislative and bureaucratic authorities. Some political scientists and sociologists believe that experts' relationships with knowledge and like-minded experts were not enough to make them powerful—they also needed expert autonomy. If experts lack autonomy and become “politicized” in their connection with elites then the power of expertise becomes subsumed into the power of elected officials, bureaucrats authorities, or interest groups (Bimber, 1996; Dargent, 2011). Unlike elite theorists, however, scholars of the politics of expertise express a lack of conviction that experts are inevitably politicized. Under the right circumstances and with the appropriate strategies, state experts gain power by freeing themselves from the instrumentalization of their knowledge by others (Centeno & Silva, 2016).

Scholars of the politics of expertise approach the issue of autonomy in two different ways, acting from different motivations. One group frames autonomy as the insulation of experts from the partisan interests of politicians such as elected and bureaucratic officials. These scholars focus on the political conditions necessary to achieve the rational dream—an ideal implicitly worth pursuing (Bimber, 1996). Joyce (2011) recounts the history of the US Congressional Budget Office and the efforts needed to maintain its independence, credibility, and ability to produce “honest numbers” over four decades. Bimber (1996) explains how one US Office of Technology Assessment director achieved greater autonomy for the office by liaising and networking with other influential actors while another director failed to garner as much respect and thus autonomy by falling back on appeals to his scientific authority. These two studies take a relatively uncritical stance towards the role of experts in governance.

A second group of scholars takes a more critical stance towards state experts. Similar to the objectives of social scientists focused on state autonomy, these scholars have been occupied with investigating the conditions and strategies both inside and outside expert institutions and organizations that did or did not permit expert autonomy (Centeno & Silva, 2016; Dargent, 2011). They argue that without autonomy, experts would be the instruments of other agents rather than agents of their particular ideas and approaches (Dargent, 2011). Expert autonomy is a concern not only because experts often serve at the discretion of and are subordinate to their hierarchical superiors but also because the autonomy of unelected knowledge-based authorities posed a danger to democratic norms and decisions (Centeno, 2010). This group does not subscribe to the rational dream or technocrat narrative that technical solutions can optimize resources and resolve social and political conflict (Centeno & Silva, 2016).

### **Rationality as Political Accomplishment**

Studies of state autonomy have effectively demonstrated the importance of autonomy as an important source of expert power. Nevertheless, they do have shortcomings. These studies principally focus on the relationship and power dynamics that exist between state experts and politicians rather than between experts. They do not discuss struggles between different expert groups and their respective rationalities *within* the state. However, a few sociologists do.

One excellent example of such a struggle is Espeland's (1998) description of conflicting rationalities between engineers and environmental analysts at the Department of Interior over the siting and construction of a dam on Native American land in the American Southwest. Another example is Reay's (2007) discussion of how economists perceived and reacted to conflicts with other professionals such as lawyers, engineers, and agronomists within the US federal bureaucracy.

By ignoring the presence of diverse groups of state experts and the roles they play in governance, studies of expert autonomy reinforce the impression among many that particular manifestations of the rational dream, such as the neoliberal or managerial state, are foregone conclusions once experts have achieved autonomy. Furthermore, by framing the autonomy of experts as a threat to democracy, these literatures bolster the rationality-politics divide among social reformers and critical theorists; expert-led governance depoliticizes liberal, representative democracy rather than represents a distinct form of politics resulting from experts' sources and strategies of power.

**Trials of Strength in Science Studies.** Unlike studies of expert autonomy, struggles between groups of scientists (or engineers) and their competing paradigms are at the center of science studies (Shapin & Schaffer, 2011; Yonay, 1998). Science studies investigates how a network of experts and their allies withstand “trials of strength” in order to stabilize their way of doing things (Latour, 1988). Science studies' concept of trials of strength represents particular scientific paradigms as political accomplishments over other paradigms. Withstanding trials of strength involves acts of power in the form of translating interests and enrolling allies in order for scientists to strengthen their intellectual projects and make them more defensible (Latour, 1988; Yonay, 1994).

Science studies presents scientists and engineers not only as knowledge-based experts but also as entrepreneurs (Latour, 1988). As entrepreneurs, they are creating something novel, and novelty can appear threatening, so it is often resisted (Garud, Hardy, & Maguire, 2007). Thus, scientists try to convince peers and outsiders—especially those that disagree with them—that their contribution or project is valid, useful, or promising. Scientific ideas, methods, and paradigms have to be promoted and defended from criticism, competition, or attack (Latour,

1988). Scientists achieve this by imposing their version of reality onto as many actors as possible through translation (Law, 2011). The steps of translation involve promoting a project (i.e., a paradigm or methodology), enrolling various human and non-human entities (e.g., natural or technological phenomena), and aligning their interests with the project by framing the problem and providing them with a new social role or identity (Cabantous & Gond, 2015). This process is similar to the one institutional scholars suggest institutional entrepreneurs engage in to make change in their organizations or fields (Battilana, Leca, & Boxenbaum, 2009). However, institutional scholars do not specify the exact strategies that *expert* entrepreneurs might employ, while science studies scholars focus more on *laboratories* than they do on institutions involved in planning and policy, such as the state (Pickering, 1992).

When translation is successful, actors who were once bystanders or opponents become enrolled as allies (Callon, 1984). Entities align themselves with the scientists' project and their network, increasing the latter's authority. The more human agents and non-human "actants" that a group of scientists can motivate to comply, the more powerful those scientists become (Law, 2011). Successful acts of translation and enrollment establish their projects as "obligatory passage points" that other actors must "pass through" to accomplish their goals (Callon, 1984). By establishing themselves as obligatory passage points, scientists force others to move along paths they have defined while barring them from other possible paths (Law, 2011). Planning theorists and political scientists also discuss how experts gain power by enrolling allies, or, what they these scholars call "building coalitions" (Baum, 1980; Benveniste, 1972). Relevant literature in science studies differs from these studies because it considers non-human entities as possible allies, details the translation process which is different from coalition building, and discusses enrolling allies specifically in order to overcome opposition by other experts.

The classic case of successful translation strategies by scientists is provided by Callon's (1984) explanation of the domestication of scallops at Saint-Brieuc Bay in Brittany, France. In Callon's account, a group of scientists was able to translate the interests of diverse actors, such as scientific colleagues, fishermen, and scallops (*pectin maximus*), in order to pursue their particular scallop cultivation strategy in a new marine environment. As a result, successful translation and enrollment led the scientists to speak authoritatively on behalf of all "enrolled entities": "At the end ... three researchers said what these entities are and want" (Callon, 1984: 210). Experts who can successfully withstand trials of strength bolster the network for their project and can overcome opposition from competing projects. This was the case of Bayesian statisticians who overcame opposition by "frequentists" in decision science and neoclassicists by institutionalists in economics (Cabantous & Gond, 2015; Yonay, 1994).

### **Experts' Methods and Power**

Studying the methods of scientists, particularly their experiments, forms a major commitment in science studies (Latour, 1993; Latour & Woolgar, 2013; Law, 2011). Science-studies scholars do not divorce the power and authority of scientists from their work in science. Earlier, I reviewed literature that considered the power of expert knowledge as a property or resource; ideas that affect interests and institutional missions; and discourse that produce identities, truth, and deviance. These are all valuable ways to understand the power of experts, but scholars have overemphasized these means at the expense of experts' practices and work. As sciences studies scholar John Law said of science: "The process of building scientific knowledge is also an active matter. It takes work and effort" (Law, 2004: 19).

When scholars examine expert power as discourse alone they, in one sense, depoliticize experts by taking their statements at face value. In describing the "dark side" of planning,

theorist Bent Flyvbjerg cautioned against relying on speech and text because experts “participate in deception ... they play games of power dressed up in technical reasoning” (Flyvbjerg, 1996: 387). Taking expert knowledge at face value divorces it from its existing political context, which includes the organizational and technical work of experts and their allies. The focus on experts’ ideas emphasizes their intellectual output but neglects the political work necessary to establish and execute their methods. According to one science studies scholar: “We cannot simply denounce knowledge as ideology; applied social science and quantification more generally is hard work” (Breslau, 1998: 109). Governmentality studies is an excellent resource for understanding the emergence of a particular *governmental program* but explains neither the role of social scientists in the development of governmental technologies such as state statistics nor whether the effects of surveillance and normalization had anything to do with the production of statistics inside state organizations (Fraser, 1981; Jessop, 2001). One could conclude, to borrow a phrase from the public administration literature, that governmentality is “governance without government” (Peters & Pierre, 1998).

Studies that examine experts’ methods introduce an important concept called “practice ontology” (Feldman & Orlikowski, 2011) to the study of state experts as a supplement to existing programs that focus on ideas and discourse. Best (2014) said that “[t]o grasp the dynamics of global governance, we therefore need to understand the production of expert knowledge as a kind of practice—a task...that is particularly suited to the insights of actor-network scholars such as Callon and Latour (Best, 2014: 24). This focus on practice recovers the expert agency lost in Foucauldian-inspired governmentality studies (Fraser, 1981; Jessop, 2001).

Expert “policy devices”, or methods for doing the work of seeing and deciding, have been neglected as a source of power for state experts—economists in particular (Hirschman &

Berman, 2014).<sup>2</sup> This is surprising since scientific method is a core tenet of rationality (Popper, 1972; Toulmin, 2009). Until recently, scholars have studied the policy devices of experts only obliquely and have often concluded that they influenced only the choice of technical matters or policy instruments, or their details (Anderson, 2008; Brint, 1990). In a set of significant early case studies Nelkin (1975) concluded that expert influence was restricted to moments without conflict and to the instruments of planning and policy rather than political values. Studies of national-income accounting have concluded that economists are more influential in debates that are technical, focused on the particulars of policy instruments, and shielded from public scrutiny (Hall, 1989). In other words, experts have an impact on the *means* but not the *ends* of planning. For social reformers, this had the effect of exacerbating the chastened dream. At the same time, these studies depict expert influence in technical matters of planning, policy, and public administration as unremarkable and politically insignificant.

Hirschman and Berman (2014) agree that experts may have the most say and influence over the technical aspects of policy, but they do not dismiss such influence as limited or apolitical. They cite policy devices as pillars of expert power with potential political effects that can sometimes extend to policy ends. Even when they do not extend this far, the effects still merit deeper empirical investigation for their “small-*p*” political effects (Hirschman & Berman, 2014). Scholars are beginning to examine the policy devices of experts and their potential political effects more closely. For example, Hood (2017) and Viscusi (2009) surveyed the efforts of the Office of Management and Budget and Environmental Protection Agency to

---

<sup>2</sup> The term *policy devices* is used in lieu of *methods* when referring to the term’s use by its original authors. The terms are meant to signify the same thing, however.



measure the statistical value of life (Hood, 2017; Viscusi, 2009). Much more terrain remains to be covered, however.

A sub-discipline of sociology, the sociology of interventions, is less concerned with expert strategies designed to achieve autonomy and thus avoid politicization, than with those designed to establish methods that do work on their behalf. Inspired in part by science studies, scholars committed to this intellectual program focus on experts' methods and how they intervene in their organizational environments and produce political effects (Asdal, Brenna, & Moser, 2007; Eyal & Buchholz, 2010). In other words, they focus on the content of what experts actually do. For example, economists usually intervene through calculative tools designed to bring about governability (Eyal & Levy, 2013).

**Opening the Black Box of (Social) Scientific Practice.** Policy devices come in many different forms such as indicators, models, matrixes, algorithms, decision trees, and (e)valuations, to name a few. A collection of literature that falls under the broad rubric of “sociology of quantification” makes sense of these diverse devices. The sociology of quantification, as defined here, includes literatures from sociology and science studies but also critical accounting studies, history, and anthropology that investigate the core functions of methodological work such as classification (Bowker & Star, 1999), standardization (Timmermans & Epstein, 2010), measurement (Power, 2004), quantification (Espeland & Stevens, 2008), calculation (Callon & Law, 2005), commensuration (Espeland & Stevens, 1998), valuation (Lamont, 2012), and ranking (Espeland & Sauder, 2007). These bodies of literature have in turn been applied in a variety of fields such as medicine, public health, public policy, and finance and economics (Berten & Leisering, 2017; Besedovsky, 2018; Maiers, 2017; Mansnerus,

2013; Mügge, 2016). These literatures open up black boxes and get inside a wide array of expert methods to understand their biases and effects.

Science-studies scholars “open the black box” to better understand scientific practice. The “black box” is a concept that appears to have originated among electric engineers and represents a device that transforms given inputs into predictable outputs in cases where the process for doing so is opaque or considered settled (Latour, 1988; MacKenzie, 2005). In practice, users of black boxes are content to disregard their internal structure and workings so long as they function as intended. Modern society is replete with black boxes from technologies such as automobiles and automated teller machines, organizations such as departments of taxation and post offices, and expertise such as surgery and sailing (MacKenzie, 2005).

One major source of the black box’s power is its ability to mask the internal workings of (social) science including a field’s paradigms or elementary theories and methods (Law, 2004; Pinch, 1992). When black boxes conceal the inner workings of science, they conceal the moral, ethical, and political content of the inputs entering the black box and choices that determine how those inputs are transformed into outputs (Eyal & Levy, 2013; Steen, 2015). This process is powerful because actors who wish to reveal these biases do what they can to open the black box while scientists do whatever they can to ensure that the black box remains closed and continues to operate reliably and opaquely (MacKenzie, 2005). When the methods of experts are successfully black boxed and adopted into governance contexts, the political and moral content of the methods’ inputs and design choices engender political effects.

Marshaling black-boxed methods as governing rationales is an important source of power for state experts. Getting inside these methods can reveal their small-*p* political effects—intended or unintended—such as what and whom experts make visible or invisible, include or

exclude, and elevate or demote (Alonso & Starr, 1989; Desrosières, 2002; Espeland & Stevens, 1998, 2008). Hirschman and Berman (2014) document how policy devices, once stabilized, had the ability to shape attention among actors. For example, the national-income accounts produced by economists have drawn attention to aggregate growth rather than the distribution of national income (Hirschman, 2016).

Methods are not limited to small-*p* political effects, however. As decisional devices, they can also have “big-*P*” political effects by influencing power relations and organizational outcomes (Nathan, 1987). For example, the World Bank’s adoption of structural adjustment models increased the influence of macroeconomists at the expense of sector specialists (Pereira, 1995). The United States Department of Labor’s adoption of net-impact analysis shifted power from street-level bureaucrats implementing labor-market reentry programs to evaluators and senior Department of Labor bureaucrats (Breslau, 1997). Regarding organizational outcomes, Carolini (2008) describes how the UN’s national-income and public-sector accounting standards concealed for-profit motivations that reduced social investments in low-income Brazilian neighborhoods. Willis (1995) describes how financial formulas contributed to the different architectural forms of early twentieth-century American skyscrapers in New York and Chicago. The influence of policy devices on power relations and organizational decisions has received relatively less attention than their impacts on how decision-makers and organizations see.

**Economic Methods.** Science-studies scholars have focused more on economists’ use of “market devices” to shape markets and market behavior than on policy devices to affect policy and planning (Hirschman & Berman, 2014). Economists affect markets whenever, over time and with repeated use, their models create the behavior, such as the characteristics of “*homo economicus*”, that they describe (MacKenzie, Muniesa, & Siu, 2007). According to this idea,

economic theories and models intervene in reality rather than simply reflect it. At the same time, economists have played an increasingly prominent role in planning and policy in the past half-century (Markoff & Montecinos, 1993). Scholars are starting to give this specific group of social scientists and their methods the attention they warrant. Economists deserve this attention because their methods employ two unique strategies: monetization and quantification. What separates economists from accountants and financial analysts in their use of monetization and quantification strategies is their fusion of these skills with the core tenets of neoclassical economic theory. State economists frequently rely on a series of “insights and practical techniques” from microeconomics (Reay, 2007). These include instinctively considering costs as well as benefits (i.e., cost-benefit analysis), incentives, market equilibrium, and opportunity cost (Reay, 2007). They also employ counterfactual thinking in which economists ask “what if?” and then try to quantify and monetize the counterfactual scenario (Abbott, 2004). I would add to this list the allegiance of economists to generalized linear reality—i.e., the general linear model and fundamental assumptions that permit its use (Abbott, 1988).

With respect to economists’ methods and according to the relevant literature, they generally narrow decision-makers’ field of vision, establish formal decision-making processes, and restructure political relations around them (Hirschman & Berman, 2014). Economists can also contribute to casting non-economic topics in economic terms and putting non-economic decision-making domains under their technical jurisdiction (Berman, 2011). Methods can help them to present themselves as official spokespersons of objectivity and rigor (Porter, 1996). They can remake the outside world in the image of an economic experiment or model (Mitchell, 2005). Methods can elevate some values while rendering others invisible (Espeland & Stevens,

1998), make economic planning and policy goals more legitimate than others, and make quantifiable ends more legitimate than qualitative ones (Centeno, 1993).

Economists can use a variety of methods, so long as they are amenable to quantification, monetization, or both. The most popular methods in contemporary developmental economics are governance indicators, decision trees, and randomized controlled trials (Akbulut, Adaman, & Madra, 2015; Rodrik, 2008). Except for the studies described in this review, limited work has been done on these specific methods, particularly in the field of development finance. The authors of an edited volume, *Governance by Indicators*, describe how indicators can displace subjective, non-quantitative data with “scientific” objective data, determine how problems are framed, and influence action, including the allocation of resources (Davis, Fisher, Kingsbury, & Merry, 2012). Generally, indicator effects include combining multiple diverse qualities or quantities into a single metric (i.e., commensuration), ranking entities along that metric, and eliciting reactions from the entities they measure and rank (Espeland & Sauder, 2007; Espeland & Stevens, 1998). Decision trees and RCTs have been analyzed in fields other than that of development. A study of decision analysis shows how decision trees disciplined the beliefs and preferences of decision-makers, visually represented a decision scenario, quantified a limited set of possible decisional alternatives, and arrived at a set of ranked options that could be implemented (Cabantous, Gond, & Johnson-Cramer, 2010). A study of impact evaluations in US domestic policy, including randomized controlled trials, showed that they can make illegitimate problems legitimate and justify contentious programs by using evaluative outcomes to argue in favor of continued funding (Breslau, 1997; Breslau, 1998). They can redistribute agency away from program recipients and toward program interventions and the bureaucrats that implement them (Breslau, 1998).

## **Conclusion**

In conclusion, this chapter reviewed the major debates over the power of experts. Orthodox and heterodox planning theorists, neo-marxists, and postdevelopmentalists contribute to the prevailing belief that experts' power in planning and development is certain and derived from their intellectual rationality. Meanwhile, others cast doubts on the power of experts' intellectual rationality. Social scientists ascribing to elite theory and a number of empirical case studies across various disciplines claimed that politicians instrumentalize experts or if they do have influence it is limited to narrow technical matters. These scholars' arguments fuel the chastened dream and asymmetrical treatment of rationality and power in planning, policy and development.

The politics of expertise qualifies the power of intellectual rationality. This literature suggests that intellectual rationality is powerful but only under conditions of "expert autonomy"—experts must do work and create the necessary conditions to liberate themselves from bureaucratic and legislative authority in order for their rationality to prevail. Science studies and the sub-discipline of the sociology of interventions likewise argue that experts must do political work to be influential. These genres are more concerned, however, with rationality as a political accomplishment—experts' power is derived from their ability to establish black-boxed methods that do work on their behalf. In other words, experts' intellectual rationality is powerful, but not, as many in planning and development believe, because of their ideas and discourse alone or, as the politics of expertise literature argues, because they achieve autonomy from politicians. Rather, experts' power is derived from their ability to overcome opposition from other experts to advance their methodological projects.

## Chapter 3: Research Approach and Methodology

### **Introduction**

This chapter will describe the research approach, design, methodology, data sources, and procedures used to answer questions around the power of economic expertise. This study faced the challenge of using research methodology to study research methods. The most appropriate approaches for such a task are phronesis and mixed methods. Phronesis focuses on planning practices and their power. Mixed methods permit an eclectic use of research designs, methods, and data sources to answer the questions at hand. Together, these philosophical approaches led to an embedded case study design using comparative case analysis and mixed methods.

The case chosen for analysis is a new fund for global development, the United States (US) Millennium Challenge Account (MCA), and the agency responsible for executing it, the US Millennium Challenge Corporation (MCC). Fifty in-depth interviews, hundreds of documents, many obtained through Freedom of Information Act (FOIA) requests, and several quantitative data sets were analyzed. The many diverse sources of empirical data allowed for multiple forms of triangulation. The unique case study design meanwhile facilitated detailed descriptions of processes and outcomes.

This chapter will first introduce the value-rational, phronetic approach. Then, it will describe the embedded case study design and mixed methods approach in detail. The primary and secondary data sources and data collection and analysis procedures supporting this approach are then presented. The chapter closes with a discussion about how this research met data quality and ethical standards.

## **Research Approach**

This dissertation investigates whether state expertise is powerful, and if so, how. It strives to answer these two questions by examining how development economic methods get stabilized within development governance and affect development agencies and programs. There are two principal philosophical approaches driving this study's methodology: "studying up" and phronesis. Studying up is an approach to researching elites (Aguiar & Schneider, 2016). As a response to ethnographers who ventured into "the field" to study marginalized populations, Laura Nader, a notable anthropologist, suggested that ethnographers instead "study up" (Nader, 1969). Studying up shifts the focus to officials and experts in positions of authority and influence in institutions such as political parties, the state, academia, corporations, the military, and non-governmental organizations (Aguiar & Schneider, 2016). I follow in this tradition by studying economic expertise in development. Much of development studies has been focused on "studying down". Too often development is focused on the recipients of development projects or the "development subject" (e.g., Yapa, 1996). Meanwhile, we know relatively less about the work of economic experts in development.

The other methodological approach is phronetic research. Phronetic research is based on Aristotle's philosophy of phronesis (Flyvbjerg, 2001). Phronesis can be translated as prudence or practical wisdom. It is considered a value rational approach rather than an instrumental rational approach. Value rationality involves reflexive analysis and discussion of values and interests (Flyvbjerg, 2004). I investigate economic expertise's role and influence in governance because I am interested and committed to asking the key questions in phronetic planning research: where are we going with such an approach to governance, who gains and loses from it, and are the outcomes desirable (Flyvbjerg, 2004)?



The issue of power is central to phronetic research. Classic Aristotelian phronesis neglected issues of power. Flyvbjerg, though, places power at the center of his approach. He goes as far as to say that “[t]here can be no adequate understanding of planning without placing the analysis of planning within the context of power” (Flyvbjerg, 2004: 292). For Flyvbjerg, the main question is not only the Weberian one of “who governs?” but also the Foucauldian question of “what ‘governmental rationalities’ are at work when those who govern govern?” (Flyvbjerg, 2004: 293). By combining a Foucauldian interpretation of power with a Weberian one, the central question becomes how power is exercised rather than merely asking who has power and why they have it.

This study’s focus on economists’ methods aligns with phronetic planning research’s emphasis on the “micropractices” of planning. Experts’ methods are the minutia of planning, or the “little things” of the world that philosopher Friedrich Nietzsche said too many had come to despise (Flyvbjerg, 2001). Phronetic planning research is decentered in its approach, taking these planning minutia as its point of departure (Flyvbjerg, 2004). Most studies of experts are concerned with their social status or big ideas (Beland & Cox, 2010; Schwendinger & Schwendinger, 1974). However, ignoring experts’ little things (like methods) can result in misleading or hyperbolic accounts divorced from the reality of experts’ daily practices (e.g., Easterly, 2014).

In effect, this study takes a practice ontology towards economists’ work (Feldman & Orlikowski, 2011). Rather than focus primarily on economists’ discursive acts, this study focuses on economists’ work to gain power and establish their methods. It also examines the work those methods do on economists’ behalf. While scholars’ focus on discourse tempts readers to continually think of things as being simpler than they are, phronetic research attempts

to avoid this problem by focusing on planning practices regardless of whether they occur inside UNDP headquarters, a local planning office, a particular plan or project, a grassroots organization, or a neighborhood (Flyvbjerg, 2004). What people actually do in planning is seen as more fundamental than what people say through discourse, text, or theory.

## **Research Design**

Value rational questions focused on the power of planning practices are best examined through case studies (Flyvbjerg, 2004). Case studies are in-depth, multidimensional investigations of social phenomena through mainly, though not exclusively, qualitative and inductive methodological approaches (Yin, 1984). They are useful for answering “how” questions and understanding issues where processes and connections are imperative but are misunderstood or incompletely comprehended (Peattie, 1983).

This study is concerned with processes of how experts introduce and stabilize economic methods in governance. Some scholars have approached the issue of economic expertise and its effects using large sample quantitative techniques such as multivariate regression models. They argue that this helps them investigate the influence of economic ideas by establishing causality and answering the “how much” question of expert influence (Chwieroth, 2007). Large sample quantitative studies are the exception to the norm. The vast majority of studies of state expertise are usually approached through case studies. This dissertation is no different. I investigate the role and power of economic expertise using a case study of United States global development finance at the turn of the 21<sup>st</sup> century.

My case study approach departs from existing approaches in important respects, however. Case studies of expertise have largely been based on historical cases or analysis of discourse. Archives are a rich source of empirical material. One can conduct a rather

straightforward discourse analysis of experts using their intellectual outputs. Because state experts are elites though, scholars often have a hard time gaining access for primary source material such as interviews (Lilleker, 2003). Each of these approaches—archives and interviews—has its advantages and limitations. Historical case studies can provide an incredibly detailed assessment of the role and work of experts. Nils Gilman's *Mandarins of the Future* is a fine example (Gilman, 2003). Yet, their holism entangles a number of experts' sources of influence such as scientific authority, social status, institutional position, and methods. The obvious problem with a discursive approach is that it assumes a power that experts may not actually have. There may be a wide gap between what experts say (or even do) and what their influence is. It is a good approach for explaining experts' intentions, but not necessarily their effects.

In contrast to historical case studies or analyses of discourse, embedded case studies involve more than one unit of analysis and are more amenable to mixed-methods approaches (Scholz & Tietje, 2002). An embedded case study approach is useful when subunits, such as particular actors, objects, or practices deserve focused attention. For example, in an organizational case study the main unit may be the organization as a whole and the subunits that receive more focused attention may be specific departments or groups of individuals (Scholz & Tietje, 2002). Cases that analyze the larger unit can miss specific phenomena's operational detail (Yin, 1984). Some important operational details are the micropractices of planning such as economists' methods.

Given the complexity of units of analysis such as experts and expertise, it is difficult to isolate the influence, if not impact, of major economic ideas and practices such as national income accounting. The difficulty is used to argue for their irrelevance (Breslau, 1998). This

supposed irrelevance could be a result of limitations in existing empirical approaches. An embedded case study design allowed me to isolate methods from other sources of expert influence such as ideas, scientific authority, social class, and institutional position. This was accomplished through focusing on methods' role in governance rather than isolating methods as a variable in a generalized linear model or similar reductionist and positivist approach (Abbott, 1988). For example, I am primarily concerned with ideas in so far as they influence economists' methods. Also, I address the issue of institutional position by treating it as an outcome of economists' attempts to gain influence and stabilize their methods rather than an explanatory variable for economists' power. Still, comparative case approaches are employed to analyze economic governance methods' impacts when they are applied in decision-making versus when they are absent. These strategies will be described in more detail below.

While I argue that case studies are the best research design to investigate economic expertise, I break from the phronetic approach's allegiance to interpretivism. For strict interpretivists, generalizing from specific contexts is both impossible and undesirable (Lin, 1998). The better approach is to explain in great detail respondents' diverse perspectives in order to explain a phenomenon within its context. Because of its focus on interpretivist accounts, phronetic research rejects attempts at generalization. It considers this the purview of *episteme*, or universal, invariable, and context-independent "scientific" knowledge (Flyvbjerg, 2004).

At the same time, a complete rejection of attempts at generalization is unwarranted. One way to bridge phronetic research and generalization is to rely on "naturalistic generalization". Naturalistic generalization is well suited for phronetic research and can contribute to externally valid and useful knowledge without trying to establish a new law or paradigm. Naturalistic

generalization is based on a partnership between readers and researchers. Readers use their intellect and tacit knowledge of their own experiences to determine whether a case study's findings can be transferred to other contexts or cases the reader is familiar with (Stake, 1983). The responsibility of each reader is to determine how much a particular case fits with his or her personal experiences or circumstances (Lincoln & Guba, 1985). The researcher's job is not to establish a representative sample and declare generalizations but provide a sufficient level of contextualized empirical information to facilitate the reader's judgment as to whether a particular case can be generalized. Such detail is often but not exclusively provided through interpretivist narratives. The researcher can also draw his or her own conclusions and develop theory so long as enough detail is provided for readers to draw their own, potentially different, conclusions (Hellström, 2008).

Embedded case studies have the advantage of maintaining the rich contextual detail of cases that permit naturalized generalization while accommodating a mixed-methods approach (Scholz & Tietje, 2002; Stake, 1983). Mixed methods research moves beyond the "incompatibility thesis", which is driven by the supposed incommensurability of quantitative, positivist and qualitative, interpretivist research paradigms (Johnson & Onwuegbuzie, 2004). It does so by mixing or combining quantitative and qualitative research techniques, methods, approaches, concepts, or language into a single study. Doing so draws on the strengths and minimizes the weaknesses of each paradigm (Johnson & Onwuegbuzie, 2004). A mixed method approach likewise focuses on what qualitative and quantitative methods share. Both approaches describe data, construct explanatory arguments from those data, take pains to ensure the validity of arguments, and speculate about why the observed outcomes happened as they did (Sechrest & Sidani, 1995).

Notably, mixed methods research is based on the philosophy of pragmatism (Johnson & Onwuegbuzie, 2004). Pragmatism focuses on identifying solutions that best help to understand or address the problem. It embraces eclecticism and pluralism in research design and methodology (Johnson & Onwuegbuzie, 2004). In contrast to studies taking a narrative, phenomenological, or ethnographic approach, pragmatist designs employ whatever sources or data best address the questions posed. As a result, the embedded case study approach could be described as more eclectic and opportunistic than narrative-based case studies.

**Description of the Embedded Case Study Design.** This study is based on an embedded case design of three paradigmatic economic governance methods within the Millennium Challenge Account (MCA) and Millennium Challenge Corporation (MCC). The focus is on three major economic methods—governance indicators, growth diagnostics, and randomized controlled trials—that comprise the embedded subunits of the broader case of the MCA and MCC. The embedded case study design permits a greater focus on these economic methods as governance rationales and sources of economic experts’ power. Each component case also employs its own unique comparative analytical approach that illuminates the methods’ effects on the distribution of development finance resources. Because the MCC adopted these methods, the case components are supported by secondary case examples from multilateral development banks (MDB) to confirm findings and give confidence about any generalizations that might be drawn.

**Embedded Case Components: Economic Methods at the MCC.** The MCA-MCC was chosen as a case study because it represents a paradigmatic case (Flyvbjerg, 2006) of changes in global development finance governance during a period of change. Three “subunits”, or embedded case components, were selected for in-depth examination and form the basis of this

study. Each of these components is a method in the development economics discipline that state economists introduced as governance rationales to the MCC. Like the overall case, these case components were selected because they are paradigmatic cases. The three methods—governance indicators, growth diagnostics, and randomized controlled trials (RCT)—were the most popular contemporary methods in both development economics and development finance during this period (and still are as of 2018). The methods’ creators work in some of the most prestigious development economics research departments in the world and are among the most widely cited and influential scholars in the field. Each of the methods has been adopted, to differing degrees and sometimes in different forms, by various development finance organizations such as the World Bank, International Fund for Agriculture, and USAID.

In addition to being paradigmatic cases, these methods are also extreme cases (Flyvbjerg, 2006)—the ideas and practices that would inform global development finance were most enthusiastically applied in the MCC. MCC adhered to governance indicator criteria more faithfully for country selection, used growth diagnostics more directly for development planning, and executed RCTs for project evaluation more often than its development finance agency counterparts. As a result, the processes and outcomes described in this study could be more pronounced than in other contexts. In this study examples and evidence are drawn, when appropriate, from secondary cases of global development finance organizations such as the World and Asian Development Banks. This triangulation corroborates findings from the primary case of MCC.

The first case component is MCC’s use of governance indicators for country selection. Governance became an important concept for development following development and international macroeconomists’ interest in the field of new institutional economics. Governance

indicators attempt to measure and quantify a wide range of institutional experiences in a country. They aggregate various expert opinions on topics such as rule of law, bureaucratic effectiveness, political stability, democratic norms, and extent of corruption in a country. Indicators then rank countries from “best” to “worst” governed according to their governance score. MCC began using governance indicators to rank and select countries for MCA grants at inception.

This case component examined how governance indicators were introduced and stabilized as part of the MCA. The process of selecting countries by indicators was compared to existing foreign policy approaches to selecting countries for aid. The MCA began with 16 indicators (as of 2016 it had 20) to provide an annual snapshot, or scorecard, of countries’ economic, governance, and human development performance. Of these 16 indicators, five were derived from the World Bank’s Worldwide Governance Indicators (WGI) database. Scoring well on the WGI indicators was critical to a country’s eligibility status. To assess the impact of governance indicators on the distribution of development finance, countries that performed well on the WGI were compared to those that did not.<sup>3</sup>

The second case component is MCC’s use of growth diagnostics to plan and design country investment programs known as “compacts”.<sup>4</sup> Growth diagnostics, called “constraints analyses” at MCC, involve a diagnostic approach to understanding bottlenecks to private investment. Rather than standardizing a particular reform program, as did the Washington

---

<sup>3</sup> Ideally the study would have compared the distribution of US bilateral economic development assistance between MCA eligible and ineligible countries since the MCA’s establishment in 2004. However, a number of confounding variables that affect the distribution of the State Department Economic Support Fund and Development Assistance Fund make such a comparison difficult. A development policy expert at the Center for Global Development advised me that such an analysis would not result in very meaningful results. The analysis in this study focused instead on governance indicators’ effect on the distribution of new MCA funds on candidate countries.

<sup>4</sup> MCC also has a “Threshold Program” that provides small grants to countries that are on the cusp of meeting MCC’s indicator eligibility criteria for larger compact grants. The Threshold Program, while subject to the economic governance methods described here, is a marginal program and not included in this study.



Consensus, growth diagnostics recommend understanding each country's unique context and various possible paths to growth through a decision-tree analysis.

MCC implemented growth diagnostics in 2009. This second case component examines how growth diagnostics were introduced and stabilized within MCC's Department of Compact Development. The technique's influence is assessed using an interrupted time-series design by comparing the development of compacts before and after growth diagnostics were introduced. Nineteen compacts were developed before 2009 when no standardized diagnostic process was involved. These nineteen compacts were compared to 14 compacts developed with a growth diagnostic to assess the impact of this method on bureaucratic power relations, development planning, and distribution of development resources.

The third component involves MCC's use of randomized controlled trials (RCTs) to evaluate investment projects. Randomized controlled trials randomly assign project participants to a control and treatment group. Because of this randomization process RCTs are considered the most rigorous form of impact evaluation, which focus on determining an intervention's effect by establishing a counterfactual. RCT proponents claim that counterfactuals created by randomization isolate the impact of a project based intervention (i.e., treatment) from other observable and unobservable confounding variables such as weather and participant motivation.

As of 2015, MCC had initiated 57 project impact evaluations (out of 144 total evaluations) of which 23 were RCTs. This third case component examined how RCTs were introduced and stabilized within MCC's Department of Accountability. I analyzed a subset of those RCTs on a particular development intervention—farmer technical assistance and training—carried out between 2005 and 2012. Among farmer training projects with RCTs, I examined one in Ghana that was evaluated using an RCT and one in Honduras where an RCT

was initiated but where the research design faltered and eventually failed during project implementation.

## **Data Sources**

**Primary Data Sources.** Empirical evidence for this study was derived from both primary and secondary sources. Primary data were derived from semi-structured interviews conducted in-person and over the phone. Semi-structured interviews employ prompts to spur free-flowing conversations and are a particularly good method for case studies focusing on elites (Beamer, 2002). The sample frame for each case component's interviewees was established based on an assessment of the key actors involved in particular processes using the internet, the professional networking site *LinkedIn*, key informants, organizational charts, and other government documents. The sample frame was not based on a statistical sample meant to achieve representativeness but instead was constructed to represent the universe of actors involved with global development finance reform and stabilization of economic methods. Interview subjects included, among others, policy advisors and officials involved in MCA and MCC's creation, economists involved in creating and conducting governance indicators, growth diagnostics and randomized trials, senior MCC officials, and program officers managing and implementing MCC activities.

When examining governance indicators the priority was to speak to members of the select group responsible for designing the MCA selection system. For growth diagnostics, I prioritized those whose tenure at MCC was long enough to have experienced compact development processes with and without growth diagnostics. For RCTs, the priority was on actors involved in bringing RCTs to MCC and designing and executing RCTs for the Honduras

and Ghana rural development projects. Across all three cases, the goal was to achieve a balance between the participation of senior officials, economists, and non-economist experts.

Interview totals for each case were drawn from this sample frame and based on time and resources available for the study and the number of individuals needed to achieve saturation on a particular line of questioning. Saturation is achieved when no new information is collected (or is expected to be collected) from respondents (Sandelowski, 1995). Individuals were prioritized for an interview based on their relevance to the study topic and questions, availability for an interview, and position within my professional network. (For sample frame and target interview quantities see

Table 3.1.)

The sample frame consisted of 95 individuals of which 59 were targeted for interviews. Contact was established with these 59 individuals and 50 were available for, agreed to, and participated in an interview resulting in a response rate of 85 percent. Interviews were conducted between June 2015 and August 2016 in Washington, DC or over the telephone. Eight interviewees who were central to the questions being asked in the study served as key informants. Their initial in-person interviews were followed up with one or more phone calls and electronic mail exchanges where issues were clarified, additional questions asked, and materials shared.

Table 3.1

*Sample Frame, Target Interviews, and Number of Respondents*

	Case 1: Governance Indicators	Case 2: Growth Diagnostics	Case 3: RCTs
Sample Frame	31	28	36
Target Interviews	19	22	18
Interviews:			
<i>Senior officials / managers</i>	6	7	2
<i>State Economists</i>	4	6	5
<i>Other State Experts</i>	1	7	4
<i>Non-state Economists</i>	1		1
<i>Other</i>	2		3
Total Interviews	15	20	15

**Secondary Data Sources.** Secondary data sources included hundreds of documents and several datasets (see Table 3.2). The main sources of secondary data included official government documents, speeches, meeting minutes, organizational charts, studies, and presentations; outputs from government consultants; think tank events' transcripts; Congressional hearings' transcripts; and development economists' publications. Other secondary data sources included newspaper articles and conference proceedings. Government documents were obtained from the MCC, Department of Treasury, Department of State, US Congress, United Nations, World Bank, Asian Development Bank, George W. Bush Presidential Library, and William J. Clinton Presidential Library.

Datasets included official MCC data on country selection and impact evaluations. Two new unique datasets were constructed from MCC documents for this study: one on MCC investments using MCC compact project descriptions, multiyear financial plans, and monitoring and evaluation plans and another on diagnostic analyses and growth constraints using MCC growth diagnostic reports. These datasets were used to answer questions about the effects of development economics governance methods.

Table 3.2

*Secondary Data Sources, Including Dataset Sources*

	Documents	Dataset Sources
Case 1: Governance Indicators	<ul style="list-style-type: none"> <li>– Transcripts of congressional hearings and testimony of the US Senate and House of Representatives</li> <li>– Archival records of the Bush Administration NSC and CEA</li> <li>– Minutes and reports of board meetings from the MCC, World Bank, and Asian Development Bank</li> <li>– Official speeches from the United Nations Monterrey Conference on Financing for Development</li> <li>– Official speeches and presentations of US government officials</li> <li>– Organizational charts of US government agencies</li> <li>– Econometric research on aid, governance, and growth</li> <li>– Reports from the Center for Global Development and Brookings Institution</li> <li>– Records and publications of the World Bank Annual Bank Conference on Development Economics</li> <li>– Newspaper articles</li> </ul>	<p>For years 2004 to 2012:</p> <ul style="list-style-type: none"> <li>– MCC candidate country reports</li> <li>– MCA eligible country reports</li> <li>– MCC Scorebooks</li> <li>– MCC country selection data</li> <li>– WGI data</li> </ul>
Case 2: Growth Diagnostics	<ul style="list-style-type: none"> <li>– Growth diagnostics and related reports for MCC countries</li> <li>– MCC guidance on conducting growth diagnostics</li> <li>– Intermediate growth diagnostic analyses</li> <li>– Report on an MCC survey of MCC’s experience with growth diagnostics</li> <li>– MCC documents on the compact development processes</li> <li>– MCC presentations on the objectives and processes of growth diagnostics</li> <li>– Guidance and evaluations of MCC’s compact development consultative process</li> <li>– Country compact agreements, including annual financial plans</li> <li>– MCC organizational charts</li> <li>– Transcripts of think tank events on MCC economic analysis</li> <li>– Scholarly literature on growth diagnostics and poverty reduction strategy processes</li> </ul>	<p>MCC investments dataset:</p> <ul style="list-style-type: none"> <li>– Compact program descriptions</li> <li>– Compact multi-year financial plans</li> </ul> <p>MCC constraints dataset:</p> <ul style="list-style-type: none"> <li>– Growth diagnostic reports</li> </ul>
Case 3: RCTs	<ul style="list-style-type: none"> <li>– RCT research design and data collection reports</li> <li>– RCT baseline and final reports</li> <li>– Field reports from project implementers</li> <li>– MCC project design reports</li> <li>– Publications on impact evaluations in development finance from think tanks</li> <li>– Transcripts from think tank events on impact evaluation</li> <li>– MCC presentations on impact evaluation and RCTs</li> <li>– MCC RCT peer review reports</li> <li>– Executive education course material from JPAL</li> <li>– Planning and implementation documents on MCC’s agricultural and rural development projects</li> <li>– Documents and presentations from MCC conferences and workshops on RCTs in agriculture and rural development</li> <li>– World Bank’s Development Impact blog</li> </ul>	<p>MCC impact evaluation database</p>

## Procedure

Several factors facilitated access to the empirical material. As a development professional and former MCC employee I had access to a wide network of individuals who served as potential interviewees. This did not, however, eliminate challenges inherent with interview-based qualitative research. People leave institutions, move away, or are unwilling to participate. Where potential interviewees were unavailable or declined interview requests, I used a snowball sampling method *via* social networking platforms such as *LinkedIn* to find replacements. *LinkedIn* shows an individual’s work and professional history, which can determine if s/he is an appropriate interviewee. It also shows how an individual is connected to a researcher or someone a researcher might know. This can facilitate introductions and interview requests. Finally, I attended meetings open to the public at MCC and related Washington, DC institutions such as The Center for Global Development and The Brookings Institution (see Table 3.3). These events, all held in Washington, DC, provided an opportunity to make contacts and ask questions.

Table 3.3

### *Public Events Attended*

Event	Date	Host
A Conversation with MCC CEO Dana J. Hyde	April 24, 2015	CGD
MCC Quarterly Public Town Hall	June 18, 2015	MCC
Global Development and the 2016 Electio	July 20, 2015	CGD
Strengthening a Bipartisan Legacy of Modernizing U.S. Foreign Assistance	July 28, 2015	Modernizing Foreign Assistance Network
MCC Quarterly Public Town Hall	September 24, 2015	MCC
Assessing the Impact of Foreign Assistance: The Role of Evaluation	March 30, 2016	Brookings Institution
What Should Tomorrow’s Aid Agencies Look Like?	June 21, 2016	CGD
Deep Dive into MCC’s Nicaragua Farmer Training Impact Evaluation	June 30, 2016	MCC

Potential interviewees were contacted by email. The email included the researcher's background, the study's purpose, how the person's contact information was obtained, a reason for why the person was being asked to participate, an explanation that their participation was voluntary, a reference to the study's Columbia University internal review board clearance, and an invitation to ask any questions about the study prior to participation. Participants were interviewed in their offices or a public place nearby. Each interview lasted for one hour on average. Respondents were informed that their identities and responses would be kept confidential. Confidentiality and anonymity were important to the study because of the potentially sensitive matters discussed. Respondents are more forthcoming and frank in their responses when their identities are kept anonymous (Leech, 2002). As a result of this confidentiality, throughout the study respondents are identified by their professional position rather than name. Interview responses were documented through detailed and copious note taking. Respondents were asked to pause or repeat responses if necessary to document potential quotes for use in the study (respondents' quotes throughout the study should be treated as approximately rather than exactly verbatim; pauses in speech and transition words such as "so", "um", "like" etc. are often excluded from quotes). Notes were transcribed into a word processing software the same day of the interview and saved in a password protected file on a password protected computer.

For the secondary data sources, I accessed documents and data through the Internet, George W. Bush Library's archives, and Freedom of Information Act requests (FOIA). Many documents, such as compact agreements between the US and partner country governments, monitoring and evaluation plans, constraints analyses, and evaluation design reports were available online. For those that were not readily available online, I obtained files by visiting the

George W. Bush Library archives in Dallas, TX and submitting twelve FOIA requests in 2015. One FOIA request each was filed with the Treasury Department, State Department, MCC, and USAID. Eight FOIA requests were filed with the George W. Bush Presidential Library, which is subject to FOIA under the Presidential Records Act. Of these twelve requests, nine resulted in the release of documents. The State Department never completed its request, the request at USAID resulted in “no records” response (i.e., the agency could not identify any records based on the criteria provided), and one request with the George W. Bush Presidential Library is still pending.

Construction of data sets took different forms for each case component. For the first case on governance indicators, MCC’s selection data for years 2004 to 2012 was downloaded from MCC’s website. This included all the selection criteria data from the original 16 indicators by country and year. This data, however, lacked country income data. Country income in the form of gross national income (GNI) per capita was available from MCC annual scorebooks. Country GNI per capita data was manually entered into the database. Data on whether a country was a candidate or eligible for MCA funds in a given year was obtained from MCC country candidate and eligibility reports and also added to existing data.

For the second case on growth diagnostics, two databases were constructed. The first database was on MCC investments for 33 compacts designed between 2005 and 2015. Compacts’ multiyear financial plans were analyzed and relevant investment data extracted and entered into a database. Dollar figures were normalized to 2015 values using inflation adjustment ratios from the Organization for Economic Cooperation and Development (OECD). Each investment was coded: 1) as a sector according to codes from the OECD’s Development Assistant Committee’s Common Reporting System for development finance organizations; 2) as



being either rural in focus, urban in focus, or focused on both (such investments were often national in scope); and 3) for whether it was an investment in policy and institutional reform. The second data set was compiled using twelve of MCC's growth diagnostic final reports. The data set included all the analyses conducted to inform the diagnostic and the diagnostic's identification of constraints. Every table, chart, or figure in the twelve reports was entered into a database and coded according to what type of analysis it represented. This included whether the analysis was descriptive or more complex (e.g., a benchmarking exercise) and whether it was a snapshot in time or longitudinal. The institutional data sources for each analysis were also entered and coded. Reports were analyzed for constraints and those constraints entered into the database and coded as either a "binding constraint" or "non-binding constraint".

For the third case on RCTs, a database of MCC's impact evaluations as of 2015 already existed. The database included every impact evaluation at MCC. Data for the evaluations included the evaluation's respective project, design and methodology, firm or consultant, start date, actual or anticipated end date, and approximate budget.

### **Data Processing**

Data processing of primary and secondary data involved three steps for each case. First, I created categories of themes and topics important for the cases such as the introduction of economic methods, stabilization of methods, opposition to methods, expert power necessary to overcome opposition to methods, and effects of economic methods. Categories included issues relevant for every case and those unique to the three case components. Next, primary data sources in the form of typed notes from interviews with participants were then analyzed with these categories in mind. Participants' statements or quotes were highlighted when they corresponded to a relevant category. The same was done with secondary data sources such as

government documents. Each secondary data source was reviewed for material that corresponded to the established analytical categories. When something relevant was found it was highlighted and coded. Finally, highlighted and coded sections from both primary and secondary sources were brought together and analyzed jointly.

Analysis of the data from the first case's country selection database focused on how performance on the World Bank's WGI and other selection indicators affected different countries' eligibility for MCA funds. Analysis of the second case's databases focused on the most frequent types of analyses conducted in the course of a growth diagnostic and the most frequent constraints, binding and nonbinding, identified in diagnostic reports. Finally, MCC's impact evaluation database demonstrated which project types and sectors were most frequently evaluated with an RCT.

### **Quality Assurance**

Data quality was assured through triangulation. Triangulation is a strategy for improving the validity of research findings by multiplying the number of data points or perspectives about phenomena under study. While verification is the default and desired goal in social inquiry, Mathison (1988) suggests that complication or contradiction is just as valuable an outcome. There are four types of triangulation: data, investigator, theory, and methodological (Denzin, 1973). Data triangulation involves using several data sources to verify, complicate, or contradict a set of research findings. In navigation and land surveying, triangulation improves confidence in any single tool or method's identification of a location. The type of triangulation employed in this study was data triangulation. Interviewees' responses to questions were triangulated with each other. There was sufficient sample size among subcategories of respondents to verify one respondent's claims with another's. Data triangulation was also employed in the analysis of

secondary data. For example, individuals' prepared remarks in speeches and congressional testimony were compared with more open responses and discussions in think tank event transcripts. Finally, findings in secondary data were used to triangulate interview responses and vice versa. There was also triangulation among cases. Findings about economic methods in MCA and MCC were triangulated with findings about similar methods in other cases from multilateral development banks.

### **Research Ethics**

Steps were taken to ensure the research process was ethical and met federal regulations for the protection of human subjects. I passed a course on Human Subjects Protection and Conflict of Interests on January 23, 2015 with Columbia University's Collaborative Institutional Training Initiative. This study underwent a complete Columbia University Internal Review Board process. The study was approved on April 29, 2015. The study was eligible for and deemed exempt from Category 2 regulatory guidelines for the protection of human subjects. As a result of this exemption, written and signed consent forms were not necessary for human subjects participating in this research project. Nevertheless, I followed the Basic Elements of Informed Consent at the outset of every interview by stating the objectives of the research, what participants could expect by participating, and that participation was voluntary.

Steps were also taken to ensure that participants' responses were protected and kept confidential. I maintained data on an external storage device used for interview data from this project. Electronic folders on that device were password protected and the password known only to the researcher. The research did not collect any personal information that was not already available in the public domain (e.g., on the Internet) through a participant's employer. Nor did the study collection any personal information of a sensitive nature. Any government documents

with individuals' personal and sensitive information was redacted by agencies' FOIA representatives prior to public release. The federal government's own internal FOIA guidelines prevent the release of individuals' private or sensitive information.

## **Conclusion**

This study answers the question of whether and how state expertise is powerful by closely examining economists' methods within the field of global development finance. It does so through a research strategy of studying up and phronesis, which both contribute to a value rational rather than instrumental rational approach to the study of experts. Both experts' power and practices are placed at the center of the inquiry.

The questions are approached through an embedded case study design of the economic methods involved in creation of US Millennium Challenge Account, a new development finance account, and how its implementing agency, the Millennium Challenge Corporation, designs and evaluates development investments. The three economic methods chosen for examination were governance indicators, growth diagnostics, and RCTs. Both the case of the MCA-MCC and these three development economic methods were selected because of their paradigmatic qualities in the global development finance field. The embedded research design was chosen for its ability to isolate, as reasonably as possible, the economic methods' effects, permit enough contextual detail to permit naturalistic generalization, and allow a mixed methods research approach.

The data sources include both primary data sources in the form of 50 interviews and secondary data sources in the form of hundreds of documents and multiple quantitative data sets. The variety of data sources allowed for robust data quality assurance through multiple forms of data triangulation. Finally, this research complied with important federal guidelines on the

protection of human subjects in research by undergoing and passing a Columbia University Internal Review Board process. The next three chapters describe the findings for each of the embedded case components derived from the research process outlined in this chapter.

## Chapter 4: The Power of Governance Indicators in Global Development Policy

### **Introduction**

This first case study examines how the Bush Administration's Millennium Challenge Account (MCA) employed the World Bank's Worldwide Governance Indicators (among others) to decide which countries to award large development finance grants, known as compacts. It first describes how World Bank economists developed governance indicators and brought them together with econometric research on "aid-effectiveness" to analyze the influence of governance on growth. This was a major catalyst for the development-aid strategy of "selectivity", or choosing countries based on their performance according to a set of defined quantitative indicators. It shows how state economists were able to employ governance and other indicators to shape the Bush Administration's strategy of selectivity that allowed senior officials to justify increases in aid funding in the face of skepticism from conservative lawmakers. The chapter goes on to describe how the economic policy community overcame opposition from the foreign policy community to establish its strategy of selectivity by indicators. After being stabilized as selection criteria, the Worldwide Governance Indicators influenced who made funding decisions and how they were made. The chapter concludes by presenting evidence of how the Worldwide Governance Indicators affected the distribution of MCA resources.

### **The Origins of Governance Indicators**

**The World Bank's Legacy of Cross-Country Research.** The World Bank's Development Research Group is considered the largest development research institution in the world (Gilbert & Vines, 2006). The World Bank's research arm, started as the Economics Department in the 1960s, went through several evolutions but always grew in size and stature. One of the most influential moments was in 1972 when former World Bank president Robert

McNamara elevated the Economics Department to the status of Vice Presidency for Research. Twenty years later, in 1992, much of the Bank's scattered research across regions, sectors, and country offices was brought together in the Development Economics Vice Presidency (Stern & Ferreira, 1993).

By the late 1990s, the Development Economics Research Group accounted for approximately three-quarters of all Bank research (Gilbert & Vines, 2006). It has at anytime about half a dozen research units named after the topics its staff researches such as Poverty, Trade, Finance & Private Sector Development, Public Services, Infrastructure, and Growth and Macroeconomics. Each unit employs, on average, fifteen social scientists, largely economists, of various ranks. In 1997 the World Bank spent US \$25 million on the Research Group's development economics research, which amounted to 2.5 percent of the Bank's entire operating budget. This money produced nearly 4,000 papers, books, and reports between 1998 and 2005 (Banerjee, Deaton, Lustig, Rogoff, & Hsu, 2006).

These financial and human resources contribute to the Bank's tremendous comparative research advantage: its ability to aggregate data for cross-country comparison. This advantage has been a feature of the Bank's research since its inception and "from the early days at the Bank, and certainly from the establishment of the research sections in the early 1970s, it was recognized that one of the Bank's strongest comparative advantages lay in its ability to assemble data and experience from many countries" (Stern & Ferreira, 1993: 547).

The World Bank's focus on cross-country data and macroeconomics in the 1980s paved the way for a handful of economists in the Research Group's Macroeconomics and Growth program to generate new datasets and conduct econometric analyses in the late 1990s. During Chief Economist Joseph Stiglitz's tenure (1997–1999), economists merged data on aid and

governance with an extensive collection of existing country data to examine novel relationships and hypotheses inspired by New Institutional Economics (NIE). They publicized their findings, well before publication in peer-reviewed journals, through the Bank's *Policy Research Working Papers* series. These papers are meant to encourage the exchange of ideas and disseminate the findings of work in progress and make an impression on academics and policymakers (Banerjee et al., 2006).

**New Institutional Economics.** An intellectual movement called New Institutional Economics transformed the way mainstream neoclassical development economists thought about economic growth. In 1993 Douglass North won the Nobel Prize in economics “for having renewed research in economic history by applying economic theory and quantitative methods in order to explain economic and institutional change” (Nobel Media, 2014). North, Oliver Williamson, and Robert Coase were the chief proponents of new institutional economics. North became the scholar most closely associated with the movement and whose work had the greatest influence on economists working on development in the 1990s. In a book that summarized much of his work he said, “Third world countries are poor because the institutional constraints define a set of payoffs to potential economic activity that do not encourage productive activity” (North, 1990: 110).

North was an economic historian who had a deep appreciation for neoclassical economic (NCE) theory. His body of work can be described as trying to get economic historians to pay closer attention to changes in institutional structures and neoclassical economists to pay closer attention to the role of institutional change in economic performance. North's early work analyzed the experiences and sources of economic growth in the United States and Europe (North, 1966; North & Thomas, 1976). Through this work, North concluded that property rights



were the most critical institution for bringing down the costs of bargaining and exchange, i.e. transaction costs. When property rights are poorly defined, say, only for a narrow group of individuals, they permit rent-seeking behavior, unjust taxation, and confiscation. Ultimately, it is the polity, or political institutions such as the state, that play a leading role in enforcing these rights (North & Thomas, 1976). But informal institutions, such as norms, are also important.

North and other institutional economist scholars wished to introduce this theory of institutions to neoclassical economics. Neoclassical economists ignored institutions because they believed they adapted naturally over time to yield a more efficient allocation of resources. This was a result of neoclassical economics' assumption of instrumental rationality and costless bargaining (North, 1994). For neoclassical economics, market exchange was a self-correcting system (North, 1994). Coase, North, and others showed that this assumption was flawed: humans operated in a world of bounded rationality—information was costly and unevenly held (Simon, 1997). This led to costly exchange that, rather than gradually correcting itself, would become path-dependent. What humans needed were institutions such as property rights to reduce uncertainty in exchange and political institutions responsible for the enforcement of these rights (North, 1990).

**The Rise of Governance Indicators.** When new institutional economics' theories were being introduced and becoming popular, no global measures of institutions that coincided with new institutional theories existed. Political scientists had been interested in institutions decades before economists but their indicators focused on questions of political instability, violence, and democracy (Williams & Siddique, 2008). These did not approximate the institutions, such as the bureaucracy and judiciary, which institutional economists considered essential for contract enforcement. Thus, economists turned to proxy indicators based on professional experts' and

regular households' subjective perceptions. These governance measures were produced largely by private organizations rather than academics (Williams & Siddique, 2008). While some of these indicators have been around for as long as political scientists' databases, the majority began to appear in the mid-1990s a few years after North's receipt of the Nobel Prize.

Organizations interested in measuring and monitoring one or several aspects of "governance"—an idea that captured multiple concepts like democracy that interested political scientists and "regulatory quality" that interested economists—assembled subjective indicators using responses to questionnaires. The vast majority of these surveys were based on the perceptions of either some externally or self-identified experts or representative households and business owners. These experts came from all professional sectors—private, public, and non-governmental—and may or may not be citizens of the countries they opined on.

Organizations commissioning these surveys included non-governmental organizations (NGO), investment risk analysis firms, multilateral development banks, and publications such as *The Economist*. In recent years some university research institutes have also begun measuring governance. The United Nations Development Program (UNDP) published a users' guide to governance indicators in 2004 that documented 33 sources with reference to an additional twenty sources (UNDP, 2004). Each indicator producer may collect information on several broad aspects of governance, such as bureaucratic effectiveness and corruption, or it may be interested in a very specific aspect of governance such as budget transparency, press freedom, or election fairness. Regardless of which organization is collecting the information, it is almost always headquartered somewhere in the Global North with an interest in measuring governance beyond its borders. Some producers concentrate on particular regions of the world, such as Africa or Latin America. Others, meanwhile, aim to measure governance globally. This variation across

producers, respondents, definitions, and regions has led to a dizzying diversity of indicators. The Organization for Economic Cooperation and Development (OECD) went as far as to call the governance indicator landscape a “jungle” (Arndt & Oman, 2006).

Some economists have attempted to tame this jungle by creating composite, or what Ravallion (2012) aptly calls “mashup”, indices of governance. Composite governance indices are indicators that aggregate multiple, diverse indicators of governance into a single measure. The primary motivation behind such exercises is to reduce measurement error: “By adding up multiple indices one may hope to get closer to that truth; in principle there can exist an aggregate index that is more informative than any of its components” (Ravallion, 2012: 3).

Among the five most widely used indicator sources, two are composite indices: Transparency International’s Corruption Perceptions Index and the World Bank’s Worldwide Governance Indicators (WGI) (Arndt & Oman, 2006). Three Research Group economists in the Macroeconomics and Growth Program—Daniel Kaufmann, Aart Kraay, and Pablo Zoido-Lobaton (“Kaufmann and Kraay”)—created the WGI. It is the most ambitious effort of its kind and has arguably become the most extensively used set of governance indicators (Malik, 2002) (producers interchangeably use the term “indicator” and “index” when referring to composite indices. The WGI will be referred to as an indicator throughout the chapter).<sup>5</sup> Like all governance composite indices the WGI aggregates multiple producers’ sources into a single indicator on a specific topic. The WGI’s topics, each comprising a separate composite indicator, include: Voice and Accountability, Political Stability and Absence of Violence, Government

---

<sup>5</sup> The difference between an indicator and index depends on what and when data is aggregated. If data is aggregated directly from survey responses, it is an indicator. If two or more indicators are aggregated together, it becomes an index. Indices are often referred to as indicators, because the underlying data has become obscured, or black-boxed. While the WGI is technically an index, throughout this chapter we will refer to the WGI as an indicator because that is how its producers refer to it.

Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. Many research economists consider these good proxy indicators for the institutions, transaction costs, and related enforcement capacities that NIE emphasizes in its work.

In the early 2000s, bilateral and multilateral development finance agencies began employing governance indicators to target their development funds. This policy strategy became known as selectivity—the concentration of development finance in fewer countries that exhibited good governance. Directing development financing selectively to countries with measurably “better” governance extended to Western bilateral and multilateral donors as a whole, including multilateral development banks, the European Union, and UN agencies such as the International Fund for Agricultural Development (Hout, 2010). The use of governance criteria represented the development finance community’s Official Development Assistance (ODA) priorities and informed private and nonprofit investment patterns originating in Development Assistance Committee (DAC) countries.<sup>6</sup>

### **Governance Indicators and Aid Effectiveness**

Neoclassical economists were able to speak with greater authority about a topic normally reserved for other experts by adding these governance indicators as a new explanatory variable for growth to their cross-country growth regressions. They could speak with greater authority about governance, a topic usually reserved for political scientists, lawyers, jurists, and philosophers, by relating governance indicators to an ontological category economists had worked decades ago to place squarely within their domain of authority—gross domestic product

---

<sup>6</sup> Established in 1960 (then as the Development Assistance Group), DAC originally had 11 members from North America, Europe, and Japan. Today it has 30 member countries, including former ODA recipients, as well as participation from the multilateral development banks. It was created as a forum for consultations among aid donors on assistance to less-developed countries. It is meant to be the “venue and voice” of the world’s major donor countries.

(GDP). Three economists affiliated with the International Monetary Fund and World Bank—Paolo Mauro, Stephen Knack, and Philip Keefer—were the first neoclassical economists to use governance indicators in cross-country growth regressions in the mid-1990s. They established the blueprint for how to use governance indicators for applied economics research.

Paolo Mauro of the IMF published “Corruption and Growth” in the *Quarterly Journal of Economics* and Stephen Knack and Philip Keefer of the World Bank’s Research Group co-published “Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures” in *Economics & Politics* (Knack & Keefer, 1995; Mauro, 1995). Both papers were published two years after North received his Nobel Prize and cite North’s book *Institutions, Institutional Change and Economic Performance* in their opening paragraphs. For instance, Knack and Keefer stated that “North (1990, p. 54) asserts, for example, that ‘the inability of societies to develop effective, low-cost enforcement of contracts is the most important source of both historical stagnation and contemporary underdevelopment in the Third World’ ” (Knack & Keefer, 1995: 207).

Cross-country growth regressions evolved alongside neoclassical economic growth theories as a way to bring theory and empirics closer together (Durlauf, 2009). The most common growth regression is a simple multiple linear regression model.<sup>7</sup> Researchers’ motivation for using the model is to test theories about what determines countries’ growth levels or rates. The researcher includes a set of independent variables known to be strongly related to growth but that are not particularly interesting. These are called control variables and usually include measures of investment, savings, and consumption (McGrattan & Schmitz, 1999). S/he

---

<sup>7</sup> A multiple linear regression attempts to model the relationship between two or more independent, or “x”, variables and a dependent, or “y”, variable by fitting a linear equation to observed data.

must also choose a set of independent variables whose relationship to growth is suggestive, hypothesized to have a significant effect, and deemed worthy of examination. These are called explanatory variables.

New institutional economics changed the growth model and led to additional growth regressions that examined the relationship between institutions and growth. Neoclassical economists used the growth regression to tie the empirical work of creating governance indicators back to the theoretical work of new institutional economics. They are principally concerned with the efficient allocation and distribution, or “flow”, of productive inputs such as capital, labor, and technology. If inputs are not flowing, it must be because there are impediments to such flows. Prior to the new institutional economics agenda, impediments were largely viewed through the lens of restrictions on market operations and deficits in human capital. The work of institutional economists pointed to an additional barrier that required amelioration—poor governance. Institutions that exacerbated transaction costs, failed to protect property rights, or failed to enforce contracts limited capital and technological investment. Like policies that restricted economic transactions, institutions could limit productivity and growth (Taylor, 2002a). If this was the case, then institutions belonged alongside economic policies and other explanatory variables in the growth regression.

Mauro and Knack and Keefer believed that investment risk analysts’ perceptions were the best available proxies for institutions: “If applied empirical economics using ‘large-*n* studies’ was to make inroads into the new institutionalism, it needed institution-based data. The discovery of country risk data compiled by private consultancies provided them” (Pistor, 2012: 171). Crucially, the three authors examined these indices’ relationship with per capita GDP. Both studies found statistically significant relationships between their indices of institutions and

various measures of GDP: “This paper has used a newly assembled data set consisting of subjective indices of bureaucratic honesty and efficiency to provide empirical evidence on the effects of corruption on economic growth. The negative association between corruption and investment, as well as growth, is significant in both a statistical and an economic sense” (Mauro, 1995: 705) and “Institutions that protect property rights are crucial to economic growth and to investment. Some of the regressions above point to effects that rival even those of education” (Knack & Keefer, 1995: 223).

It’s not clear that North would have agreed with this use of his ideas. North warned that changing institutional circumstances in what he called the Third World was no easy task. Transferring the formal rules of successful western market economies to the developing world was no panacea for solving poor economic performance and even when formal rules were changed informal norms would follow only gradually, if at all (North, 1995). If governance indicators black-boxed some of North’s more complex and nuanced arguments around institutions, then including these indicators as determinants of growth in growth regression models translated and instrumentalized North’s ideas for development policy:

Beyond the historical case studies about economic growth and development in the United States, which North offered, there was little hard empirical evidence to support the notion that the asserted relation between private rights, markets, and growth held true around the world. The new applied institutional economics, which was formed by Mauro, Knack and Keefer, and others, provided this evidence...Critically, the indicators were not simply used in their original form; they were transposed from descriptive indicators about investment conditions into normative tools for policy reforms (Pistor, 2012: 171).

Steps from where Kaufmann and Kraay were creating the WGI and Knack and Keefer were comparing institutional indices and growth, another pair of World Bank Research Group economists was bridging governance indices with research on aid effectiveness. Craig Burnside and David Dollar published two Policy Research Working Papers that would fundamentally

depend development and aid policy. In their first Working Paper, “Aid, Policies, and Growth” (1997), they introduced economic policies into the growth equation by interacting, or multiplying, a policy index (meant to represent a country’s “economic policy environment”), with a variable for foreign aid (aid as a percentage of a country’s GDP).<sup>8</sup> They found that aid has a positive, causal relationship with growth but only in a “good” policy environment, i.e. where a country’s budget surplus and inflation rate are low and trade openness is high. In a second Working Paper published the following year, “Aid, the Incentive Regime, and Poverty Reduction” (1998), they expanded on earlier work by incorporating Knack and Keefer’s Institutional Country Risk Guide index (1995) into their policy index (Burnside & Dollar, 1998). The new index that included measures of property rights, corruption, and bureaucratic quality was called an “index of economic management”.

Later that year David Dollar brought these two empirical exercises together in a World Bank Policy Research Report, *Assessing Aid: What Works, What Doesn’t, and Why* (World Bank, 1998). In one of the report’s appendices, Dollar indicated that the index of economic management had a positive, statistically significant, and causal relationship with per capita GDP.<sup>9</sup> The report summarized the findings of that analysis and Burnside and Dollar’s broader research agenda thus: “Aid effectiveness largely depends on the institutions and policies of recipient countries. Where economic management is sound, aid leads to higher private

---

<sup>8</sup> The index is the combination of a country’s budget surplus, inflation rate, and trade openness. The authors weighted the individual indicators’ values in the index according to their correlation with growth. The index can be interpreted as a country’s predicted growth rate given its budget surplus, inflation rate, and level of trade openness. Trade openness itself is an index generated by economists Jeffrey Sachs and Andrew Warner. Warner was, incidentally, the first economist hired by MCC.

<sup>9</sup> Burnside and Dollar’s use of an instrumental variable to establish causality was controversial and a point of contention. Development economists grew increasingly weary of instrumental variables as the decade progressed. This led, in part, to the rise of randomization as an objective, untainted, and supposedly unproblematic instrument for development economics research.



investment, more rapid growth, lower infant mortality, and a faster decline in poverty. Where economic management is poor, aid has little effect on development” (World Bank, 1998: 28). It also pivoted from careful and qualified research conclusions to forceful policy recommendations: “Clearly development assistance needs to be more concentrated on where it can be most effective in reducing poverty. It needs to take more account of the environment in which it is placed. Finance is most effective in reducing poverty in those countries that have both mass poverty and a good policy and institutional environment” (World Bank, 1998: 44).

### **Stabilizing Governance Indicators: Making Aid Selective**

Burnside and Dollar’s research became very influential in the development finance community (Interviewee 1A, personal communication, June 1, 2015). It was a major catalyst for the development-aid strategy of selectivity or choosing countries based on their performance according to a set of defined quantitative criteria. Development agencies immediately began citing Burnside and Dollar’s working papers. For example, the Canadian International Development Agency put out a draft policy paper in June 2001 that said World Bank researchers provided compelling evidence that good governance and a sound policy environment are the most important determinants of aid effectiveness (Easterly, 2003). Bilateral and multilateral development agencies eventually turned this praise into policy (Hout, 2004, 2010). The Burnside and Dollar studies became a major catalyst for selectivity and the “intellectual support” for the Bush Administration’s Millennium Challenge Account (MCA) (Interviewee 1C, personal communication, June 16, 2015). Senior administration officials routinely referenced Burnside and Dollar’s research conclusions and The White House described the new initiative in the following way: “The MCA is selective, targeting those countries that ‘rule justly, invest in the health and education of their people, and encourage economic freedom’. By selecting on those

countries that have adopted policies that encourage growth, MCA assistance will more likely result in successful sustainable economic development” (Foreign Affairs Budget, 2003: 74).

Selectivity was adopted because it dovetailed with the Administration’s broader global economic growth agenda and philosophy of metrics-based management. It simultaneously addressed two problems: increasing aid, which was a key demand of the global community, and accounting for aid’s contribution to growth, which was a conservative governing philosophy.

Administration economic policy makers could support good governance because it represented a set of preconditions to enable market-driven development (Khan, 2007). The Bush Administration was very worried about a slowdown in global growth following the September 11<sup>th</sup> terrorist attacks. There was a push to “come up with a prescription for growth” among National Security Council (NSC) and US Treasury officials (Interviewee 1J, personal communication, August 5, 2015). This strategy was decidedly market-driven. John Taylor, Treasury’s Undersecretary of International Economic Affairs, was focused on free trade, reducing barriers to investment, and human capital. The Administration felt that these could ameliorate a drop in global growth from the terror threat. The strategy was to “ignite a new era of global economic growth through free markets and free trade” (Bush, 2002b: 17). The Bush Administration’s National Security Strategy released in September 2002 stated that “[w]e will use our economic engagement with other countries to underscore the benefits of policies that generate higher productivity and sustained economic growth...The United States and other developed countries should set an ambitious and specific target: to double the size of the world’s poorest economies within a decade” (Bush, 2002b: 21)

Administration leaders had three principles that they believed increased growth in lower and lower–middle income countries. These were “ruling justly”, “investing in people”, and

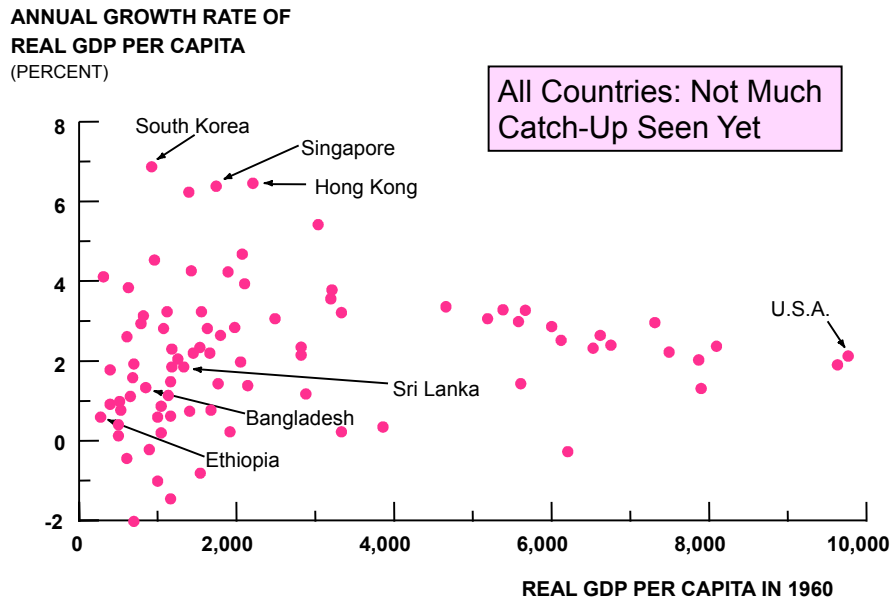
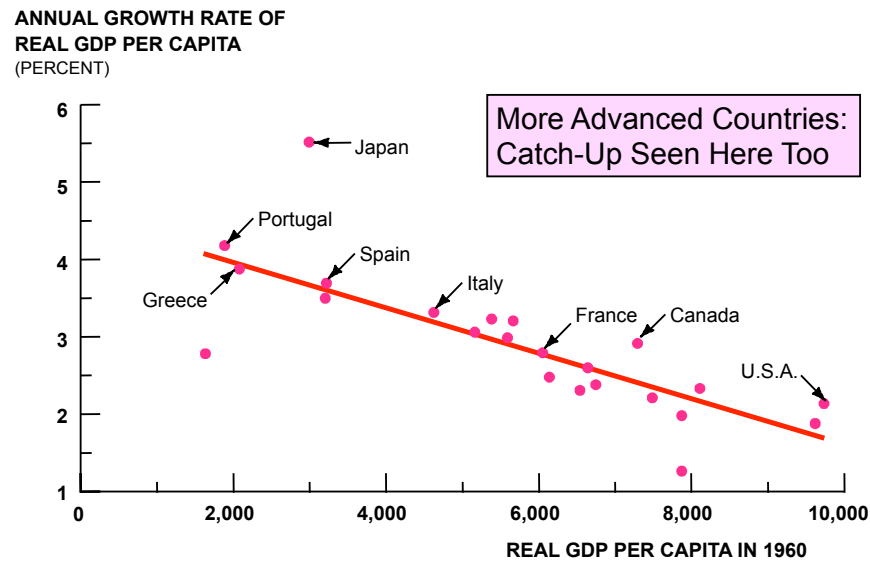
fostering “economic freedom”. “Ruling justly” refers to good governance, or rule of law, contract enforcement, and control of corruption. “Investing in people” refers to human capital development, specifically through education and health. “Economic freedom” means unhindered trade in goods, services, and new technologies. These categories corresponded to what economists believed were the principal barriers to the accumulation of physical capital, human capital, and technology.<sup>10</sup>

Taken together, the administration believed that these principles and programs would restore the flow of investment to low- and middle-income countries (Bush, Hubbard, & Kroszner, 2003). This investment would translate into higher productivity and better paying jobs. This causal narrative foretold the orthodox economists’ Holy Grail of “catch-up” where lower-income countries attain the productivity and per capita GDP levels of wealthier countries over time. This pattern has been observed in Europe and North America but proven stubbornly unattainable globally (see **Figure 4.1**). The principles of economic freedom and human capital for generating investment, productivity, and growth were well established in economic thought and policy but did not resolve the catch-up problem. At the turn of the century good governance emerged as the newest potential antidote to global disequilibrium. According to the chairman of the Council of Economic Advisors, Glenn Hubbard, this required the development community to “[t]ake governance and growth seriously” (Hubbard, 2002).

---

<sup>10</sup> According to thinking among Bush administration economic officials, poor governance creates disincentives to invest, start new firms, or expand existing businesses. This negatively impacts capital formation and entrepreneurship. Workers without adequate education do not have the requisite skills to work in high productivity jobs or adopt technology and unhealthy workers cannot work as often or as hard. Lack of openness to international trade, monopolistic state marketing boards, and excessive regulations create disincentives for the private sector to invest and innovate (Bush, Hubbard, & Kroszner, 2003; Taylor, 2002c).

**Figure 4.1.** The Puzzle of Global Disequilibrium According to the US Treasury



Source: Taylor (2002c), *New Policies for Economic Development*. presentation given at the Annual Bank Conference on Development Economics in Washington, DC on April 30, 2002.

It was unclear in the early stages of the Administration, however, what role aid would play in this growth agenda. There was considerable suspicion around aid, especially in a

conservative Administration and Congress. A director in the NSC's International Economics division said he was "working with a bunch of Republicans who thought development was money down a rat hole" (Interviewee 1F, personal communication, July 2, 2015). According to a former USAID administrator, "There was not much support among conservatives anywhere unless dynamics of the aid system were changed" (Interviewee 1M, personal communication, August 28, 2015). At the same time, calls for increased aid were growing globally. Countries' leaders adopted the Millennium Declaration at the United Nations Millennium Summit held in September 2000 that included the series of time-bound targets known as the Millennium Development Goals (MDG). A key concern of the Millennium Declaration was financing the MDGs: "We are concerned about the obstacles developing countries face in mobilizing the resources needed to finance their sustained development ... We call on the industrialized countries to grant more generous development assistance" (United Nations, 2000: 4).

After The Millennium Summit, Senegal's president Abdoulaye Wade led a delegation of African presidents from Algeria, Mali, Nigeria, and South Africa to the 27<sup>th</sup> G8 summit in Genoa, Italy to present the New Africa Initiative, which called for "a new relationship of partnership between Africa and the international community, especially the highly industrialised countries, to overcome the development chasm that has widened over centuries of unequal relations" (New African Initiative, 2001: 2).<sup>11</sup> At the G8, "After three days of discussions, everyone, including the American President George Bush, was aware of the chasm between the North and the South, between the rich of the G8 and Africa" (Harsch, 2001: 4). Part of the New

---

<sup>11</sup> The New Africa Initiative would later be called the New Partnership for Africa's Development, or NEPAD.

African Initiative strategy called for the mobilization of development finance from private and public sources, including official development assistance (New African Initiative, 2001).<sup>12</sup>

Following Genoa, G8 leaders called on the U.S. to meet its commitment to 0.7 percent of annual GDP for development assistance.<sup>13</sup> Civil society organizations were also raising their voices. Bono, the Irish rock star and lead singer of U2, along with Bobby Shriver and Jamie Drummond, established an advocacy NGO called DATA (Debt, AIDS, Trade, Africa) to lobby for greater debt relief and aid for Africa. Despite these calls from UN, African, European, and NGO leaders and President Bush's own evolution on the issue, Bush and other conservatives in both the executive and legislative branches were still skeptical about the US committing greater resources to ODA: "The President believes that the development debate has focused too much on inputs on levels of aid—arbitrary levels of aid from the developed countries, and not enough on ensuring tangible outcomes" (White House, 2002: 1).

This gulf between global demands and domestic apprehensions presented a problem in search of a solution. How would the Bush Administration contribute to these global efforts that were gaining momentum and reaching consensus if the broader perception among conservative Republicans was that aid was failing? Gary Edson, Bush's Deputy National Security Advisor for International Economics and a close confidant, was familiar with "some World Bank economists' research where countries that focus on good governance and their own efforts experience greater growth" (Interviewee 1K, personal communication, August 11, 2015). White House officials started looking for a way to tie aid into their overall economic policy agenda. If

---

<sup>12</sup> The OECD defines ODA as funds that are: 1) provided by official agencies, including state and local governments, or by their executive agencies; 2) administered with the promotion of the economic development and welfare of developing countries as their main objective; and 3) concessional in character and convey a grant element of at least 25 per cent (calculated at a rate of discount of 10 per cent).

<sup>13</sup> The United States is the largest bilateral donor in absolute terms but falls below the average commitment of the G8 in relative terms.

aid, as a source of capital, could be coupled with good policies and institutions, it could contribute to the Administration's global growth objectives.

**Creating Standards for Selectivity.** Gary Edson called Alan Larson, Undersecretary of State for Economic, Business, and Agricultural Affairs, one Sunday morning in September 2001 to discuss the President's interest in "doing something big" in terms of development (Interviewee 1D, personal communication, 2015). Larson had recommended that Edson speak to Steven Radelet, Deputy Assistant Secretary of the Treasury for Asia, the Americas, and Africa, who was known to be "a development guy".<sup>14</sup> Edson asked Radelet during a White House breakfast meeting in December 2001 how these well-governed countries could be picked. That was when the idea of using indicators to select countries for MCA assistance began. As of then, selectivity had no "flesh on the bones" (Interviewee 1E, personal communication, June 28, 2015). About two months later, in his first public comments on the MCA at the Inter-American Development Bank (IADB) headquarters in Washington, DC, President Bush said, "I'm here today to announce a major new commitment by the United States to bring hope and opportunity to the world's poorest people...Today, I call for a new compact for global development, defined by new accountability for both rich and poor nations alike...These funds will go into a new Millennium Challenge Account" (Bush, 2002a). Bush went on to outline the MCA's approach to selectivity, "I've directed Secretary [Colin] Powell [State] and Secretary [Paul] O'Neill [Treasury] to reach out to the world community, to develop a set of clear and concrete and objective criteria for measuring progress. And under the Millennium Challenge Account, we will apply these criteria rigorously and fairly" (Bush, 2002a).

---

<sup>14</sup> Radelet's reputation as a development expert came from 15 years at the Harvard Institute for International Development where he held several senior positions until Larry Summers brought him into the Clinton Administration in 2000 (Interviewee 1E, personal communication, June 28, 2015).

Later that month, over fifty world leaders (including Bush), two hundred ministers, and myriad development experts, civil society representatives, and business leaders gathered at the first ever United Nations International Conference on Financing for Development in Monterrey, Mexico. The Center for Global Development attended the Monterrey Conference as an accredited NGO and civil society member. On the “sidelines” of the Monterrey Conference, CGD held a workshop attended by InterAction, a development and aid advocacy alliance of over 150 NGOs, the Carter Center, African and Latin American development leaders, and Hilde Johnson, Norway’s Minister of International Development. This was the first of a series of five workshops on the MCA that the CGD would hold between March and September of 2002. The Center for Global Development, meanwhile, was advocating for “eligibility standards” to select these well governed countries. Established only a few months before the Monterrey conference, this development think tank was bringing together a diverse group of development experts to advise the Bush administration on the MCA selection system’s design.

Two more workshops were held at CGD offices in Washington, DC on April 5<sup>th</sup> and April 16<sup>th</sup>. George Soros, a wealthy philanthropist and democracy advocate, and Jeffrey Sachs, a prominent development economist, were keynote speakers. The sessions were invitation-only and targeted to “senior specialists from the executive branch, Congress, the NGO community, the research community, and international financial institutions” (CGD, 2002: 1). The April workshops’ objective was, according to the invite, “to help policymakers think through the analytical issues relevant to designing and implementing the performance benchmarks to which disbursements from the Millennium Challenge Account will be linked” (CGD, 2002: 1). Some questions the sessions posed were: “What should be the performance benchmarks in the three principal areas identified by the President: good governance, investment in human capital, and



economic freedom? Should anything else be benchmarked, and what priority should each area have relative to the others?” (CGD, 2002: 1).

The Monterrey Conference and April workshops culminated, in part, in a report released on May 22, 2002 titled *Guiding Principles for Design and Implementation of the MCA* by Nancy Birdsall, Sarah Lucas, and Sonal Shah who all played a role in the Washington, DC and Monterrey, Mexico sessions. The report laid out a hypothetical proposal for the MCA’s design and implementation, particularly how to determine countries’ eligibility for funding, based on seven “principles” (Birdsall, Lucas, & Shah, 2002). While the actual MCA design departed in several ways from CGD’s original proposal, it nonetheless showed the think tank’s early level of engagement and leadership on the MCA. The CGD was not just engaged in an intellectual exercise but building a network in support of selectivity through indicators. According to a former USAID administrator, “Think tanks are the intellectuals that have a lot of influence over policymaking, and in any administration, left or right, think tanks are influential” (Interviewee 1M, personal communication, August 28, 2015).

**Solving the Aid Effectiveness Problem with Governance Indicators.** Steven Radelet, a CGD affiliate, served as a key link between World Bank economic researchers, NSC officials concerned with growth, CGD experts focused on selection standards, and state economists at Treasury. He was central to introducing governance indicators into the Treasury department’s efforts to design the MCA selection system. Radelet transitioned from his position as Deputy Assistant Secretary at Treasury to Senior Fellow at the CGD in July 2002 while the think tank was conducting its MCA workshops. According to one State Department official, “Radelet was a huge force; the intellectual underpinning of the MCA model and selection approach”

(Interviewee 1A, personal communication, June 1, 2015).<sup>15</sup> Radelet reached out to his personal network, which included Research Group economists David Dollar, Daniel Kaufmann, and Aart Kraay, to brainstorm about operationalizing selectivity (Interviewee 1E, personal communication, June 28, 2015). He referred to Kaufmann and Kraay, the WGI creators, as “my friends at the World Bank” (*The Millennium Challenge Account: A New Way to Aid*, 2003: 44). Daniel Kaufmann and Aart Kraay met with the Bush administration team at least three times to discuss the WGI and their suitability for selecting countries for assistance: “When we started to narrow in on specific indicators, that’s when we would sit down with people like Dani Kaufmann and his team...we would sit down with them to dig into the actual construction and coverage of the indicators themselves” (Interviewee 1J, personal communication, August 5, 2015). Radelet’s ties with Kaufmann and Kraay brought the WGI under early and serious consideration as a set of indicators for selecting countries: “Burnside and Dollar provided the intellectual support for the [MCA] initiative and Kaufmann and Kraay provided the indicators to make it happen, the tools to operationalize something like conditioning aid on a policy environment” (Interviewee 1C, personal communication, June 16, 2015).

The Administration and Bush’s directors in the Office of Management and Budget (OMB) in particular were highly receptive to governing MCA country selection by indicators (Interviewee 1A, personal communication, June 1, 2015). Designing government programs to employ private sector–inspired techniques came naturally to Bush, the United States’ first president with a master’s degree in business administration (Pfiffner, 2007). “Metrics permeated through everything we did,” noted one senior administration official (Interviewee 1H, personal communication, July 19, 2015). For example, in the first week of his presidency, Bush proposed

---

<sup>15</sup> In fact, Radelet literally wrote the book on the MCA while a CGD Senior Fellow (see Radelet, 2003)

the *No Child Left Behind Act*. *No Child Left Behind* aimed to address educational “failure” through much greater federal control over local schools and an accountability regime rooted in standardized testing (Mehta, 2013). This thinking carried into the Administration’s development assistance programs because of the high level of agreement in the Administration around results-based management. The idea that development assistance should produce measureable results was one of Treasury Undersecretary John Taylor’s ten departmental “rules” and was a widely accepted administrative principle by other economic policy officials in the NSC and CEA. Taylor was fond of saying, “What gets measured gets done” (Taylor, 2005). Selectivity allowed a simultaneous tracking of governance and growth through indicators to judge progress and, more importantly, make the claim to conservative skeptics in Congress that aid was being delivered on the basis of evidence and accountability.

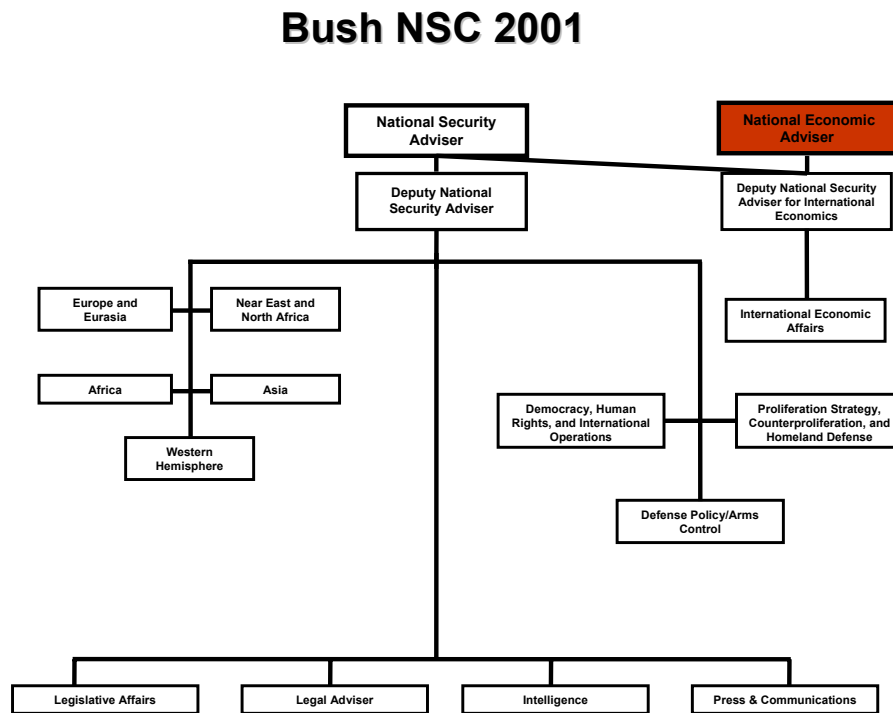
### **Conflicting Rationalities in Country Selection**

Senior NSC and White House officials consented to a strategy of aid selectivity because it resolved their conundrum of increasing aid while accounting for aid’s contribution to growth. They left the details up to development experts such as CGD fellows and Treasury economist Steven Radelet. They were the main forces behind the definition of MCA selection criteria, which included proposing indicators for country selection. The use of governance indicators like the WGI, in addition to other indicators, to distribute aid funding was a novel, (neo)liberal internationalist approach to bilateral aid. We now turn our attention to how this approach was at odds with the rationality of political realists, who have historically distributed foreign aid through a geopolitical rationale, in the US federal government’s interagency process.

**The Interagency Process: Origins of the Millennium Challenge Account.** Once selectivity through governance indicators was agreed upon as an approach, the task of selecting

these indicators and designing the selection system fell to an interagency working group (“Interagency Group”). The MCA and its implementing agency, the Millennium Challenge Corporation (MCC), were created within the Bush Administration’s NSC interagency policy process (see **Figure 4.2**). The Deputy National Security Advisor (NSA) for International Economics, Gary Edson, chaired the process (President George W. Bush created the Deputy NSA as a new position in his 2001 National Security Presidential Directive). Two statutory and two non-statutory agencies participated in the MCA interagency process: the departments of State and Treasury and the Department of Commerce and United States Agency for International Development (USAID), respectively.

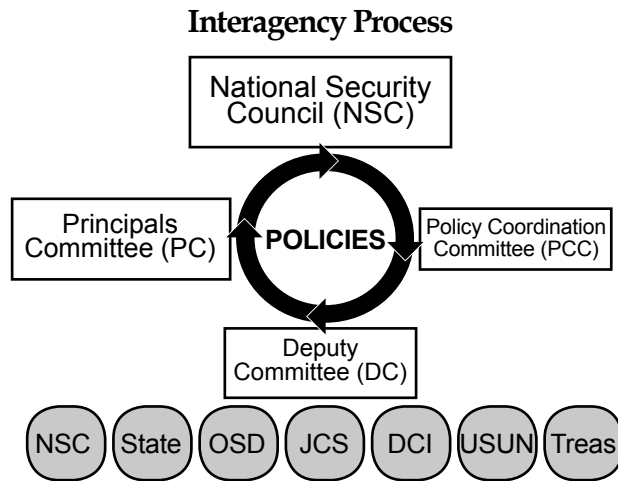
**Figure 4.2.** Organizational Structure of Bush’s National Security Council



Source: University of Maryland, (1999). *National Security Council Project*. Retrieved April 23, 2018, from <http://www.cissm.umd.edu/projects/national-security-council-project>

There are three main levels to the interagency process in descending hierarchy: the Principals Committee, Deputies Committee, and Policy Coordinating Committee (see **Figure 4.3**). The Principals Committee is the highest-level interagency forum for consideration of policy issues affecting national security. The NSC Principals Committee's regular attendees included the Secretary of State, the Secretary of the Treasury, the Secretary of Defense, the Chief of Staff to the President, and the Assistant to the President for National Security Affairs (White House, 2001). The NSC Deputies Committee included deputy secretary level officials such as the Deputy Secretary of State who monitored the work of the interagency process, did crisis management, and when necessary, pushed unresolved issues to the Principals Committee for resolution. The NSC Policy Coordinating Committees managed the development and implementation of national security policies by multiple agencies of the United States Government (USG). Each NSC Policy Coordinating Committee included representatives from the executive departments, offices, and agencies represented in the Deputies Committee (White House, 2001). The Policy Coordinating Committees were the primary fora for day-to-day interagency coordination and provided policy analysis for consideration by the more senior committees of the NSC system. President Bush established six regional and eleven functional Policy Coordinating Committees. The International Finance Policy Coordinating Committee managed the MCA design process.

**Figure 4.3.** The NSC Interagency Participants and Policy Making Process



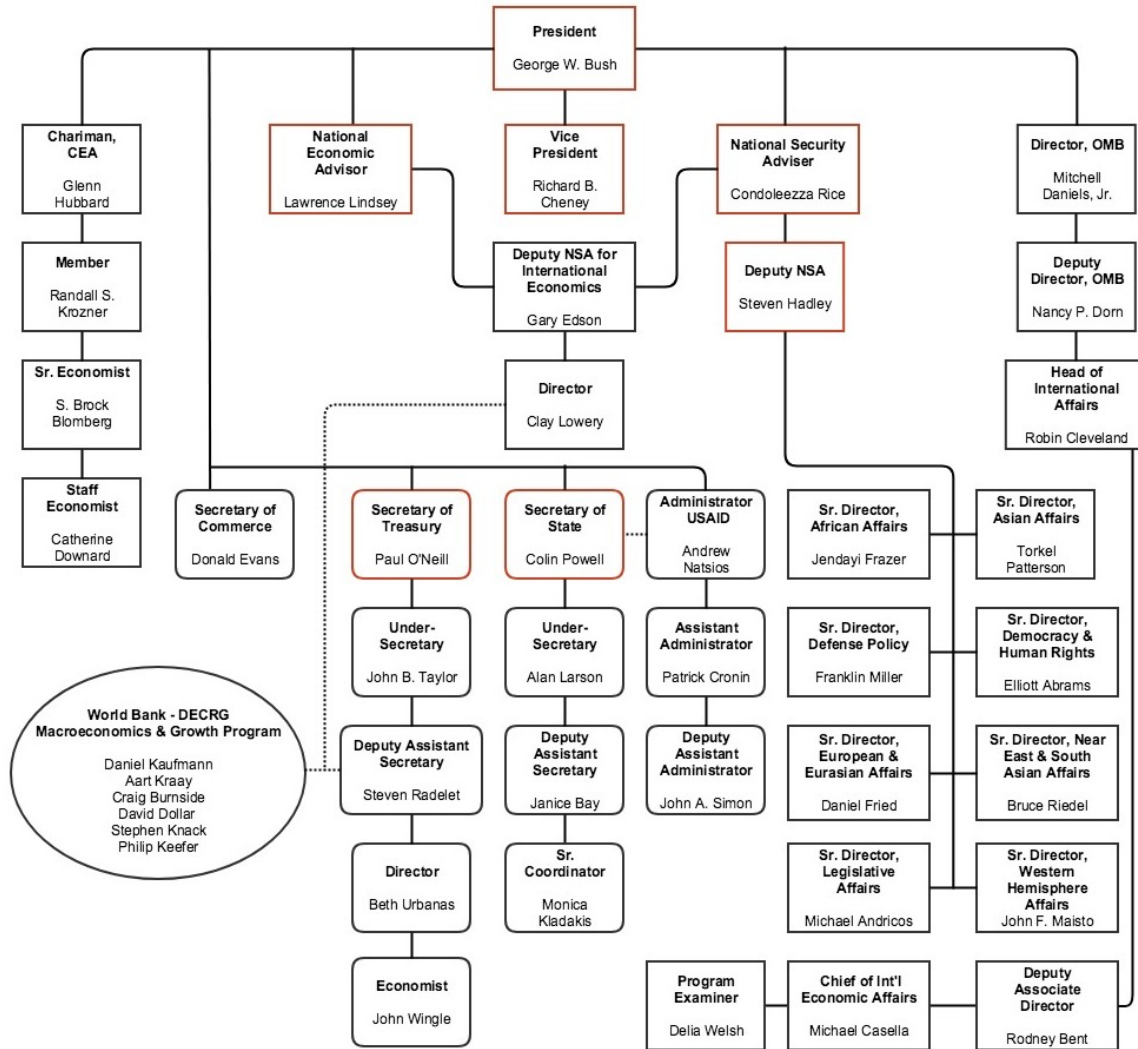
*Source:* Marcella (2008). *Affairs of State: The Interagency and National Security*. The Army War College Strategic Studies Institute. Retrieved from <http://www.dtic.mil/docs/citations/ADA491602>.

*Notes:* NSC: National Security Council; State: Department of State; OSD: Office of the Secretary of Defense; JCS: Joint Chiefs of Staff; DCI: Director of Central Intelligence; USUN: United States Mission at United Nations; Treas: Department of Treasury.

Key members of the NSC Principals Committee for the MCA were Assistant to the President for National Security Affairs Condoleezza Rice, Secretary of State Colin Powell, Treasury Secretary Paul O’Neill, and Commerce Secretary Donald Evans. Key members of the NSC Deputies Committee for the MCA were Gary Edson, Alan Larson, John Taylor, and USAID Administrator Andrew Natsios (see **Figure 4.4**). Officials and experts were “detailed”, or asked to serve a temporary rotational assignment, from their respective agencies to join one of three working groups located within the International Finance Policy Coordinating Committee. Clay Lowery, Director for NSC International Economics, chaired the Performance Indicators Working Group (“Working Group”). Senior officials in the Office of Management and Budget’s (OMB) International Affairs Office chaired the MCA Administrative Structures group that determined how the MCA would be administered. A group on MCA Legislative Strategy was also formed later on in 2003 to facilitate passage of the *Millennium Challenge Act*. Agency

personnel worked together with multiple White House Executive Offices including the NSC, CEA, and OMB (see **Figure 4.4**). In all, approximately 20 key executive branch officials were involved in designing the MCA selection process, passing legislation, and establishing the independent corporation (MCC) that would eventually manage the funds. At times, these three groups consulted with civil society groups, think tanks, and academics. The whole process lasted two years (January 2002–December 2003). It culminated with passage of the Millennium Challenge Act in January of 2004.

**Figure 4.4.** Organizational Chart of Key Bush Administration White House and Departmental Personnel Involved in the MCA's Creation



*Source:* Constructed by author based on interviews and multiple primary and secondary source documents.

*Notes:* Rectangles with hard edges represent White House staff. Rectangles with soft edges represent US executive branch departments and agencies. Ellipses represent non-USG agencies.

**The Interagency as an Arena: Conflicting Rationales for Aid.** The history of post-WWII US foreign assistance has been a marriage of geopolitics and geoeconomics. Geopolitics is about the effects of geographical location and other geographical features on the foreign policy of a state and its relations with other states. For example, a foreign policy rooted in geopolitics



may assess a state's strategic global location, natural resources, and access to waterways (Mamadouh, 1998). Geopolitics can involve economic components, such as bilateral trade agreements, sanctions, and the geostrategic provision of aid for otherwise liberal and moral goals such as social, political, and economic development. This is referred to as geoeconomics, or the use of economic instruments to achieve geopolitical goals. Blackwill and Harris (2016a) aptly refer to geoeconomics as "war by other means".

Geopolitics and geoeconomics are forms of political realism. Political realists believe that human nature is governed by egoistic self-interest, states are the most important actors in international relations, and the international arena is characterized by anarchy rather than the rule of law (Morgenthau, 1963). As a result, states are principally concerned with the pursuit of their own self-interests, which requires the attainment and projection of power. Geopolitical realism was the primary US foreign policy orientation during the Cold War. Bilateral relations were evaluated on states' allegiances to one or the other superpower and their location in "spheres of influence" such as Eastern Europe (Soviet Union) and Central America (US). The Marshall Plan was organized around economic and technical assistance but had the broad strategic objective of territorially containing the Soviet Union's sphere of influence. USAID's bureaucracy was modeled on the State department and had four world regional bureaus. Its work was often planned or assessed foremost through a geopolitical lens and intended to make foreign assistance a strategic tool to achieve foreign-policy objectives even if most actual development assistance was built on economic presumptions about growth for development (Essex, 2013). Only as the Cold War wound down in the late 1980s did universal economic rationales of liberalization and market-oriented reforms begin to take center stage in USAID's programming and US foreign assistance (Essex, 2013).

In the MCA interagency process, many Department of State officials, USAID officials, and NSC regional department members—the foreign policy community—approached MCA country selection through a realist and geopolitical lens. Neoclassical economists’ rationales for aid, meanwhile, fit squarely within the liberal internationalist school of thought. Aid given to achieve the universal objective of growth through good governance is a strand of liberal internationalist thinking. Liberal internationalism is fundamentally a vision of an open, rule-based system in which states trade and cooperate to achieve mutual gains (Ikenberry, 2009). Within this framework, foreign aid is a set of programmatic measures designed to enhance the socioeconomic and political development of recipient countries. When the flow of economic aid from the developed to the developing countries can enhance world peace and prosperity, aid is often considered part of the idealist school of liberal internationalism where ethical and moral arguments are often mobilized to justify aid (Hattori, 2003; Pankaj, 2005). The trend that began in the early 1980s of using Western economic aid to enroll countries in the international liberal order through free and stable markets in the name of growth is referred to, often pejoratively, as neoliberal (Ikenberry, 2011).

Neoclassical economists generally disapprove of realist approaches to giving aid: “The standard-bearers of economic thought during the nineteenth and early twentieth centuries had little problem using economics as an instrument of state power, whereas their neoclassical successors thought that markets were best kept free from geopolitical interference” (Blackwill & Harris, 2016b: 107). To neoclassical economists, choosing countries based on geopolitical rationales amounts to “selection bias” (Interviewee 1L, personal communication, August 12, 2015). They much preferred selectivity through indicators, which has been called a “purist” approach to ODA for its focus on pinpointing funds to achieve a narrow objective—growth:

“The MCA’s allocation criteria, based solely on economic performance and governance, would be the closest to a development purist’s blueprint for aid that the United States has ever attempted. In many respects, the MCA is precisely the sort of fund that development advocates had hoped would emerge as the Cold War wound down” (Brainard, 2003: 155). This type of approach to aid is economistic and aspatial. At a State Department conference on sustainable development, CEA chairman Glenn Hubbard put it like this: “If the MCA is to achieve its laudable objective of triggering growth by encouraging good governance... Good performance in the poor nations must be recognized regardless of the region or country in which it is achieved” (Hubbard, 2002: 7).

Nevertheless, the foreign policy community can make a case for geopolitically motivated aid and an established social science discipline, political science, is dedicated to understanding the determinants of aid allocations rather than aid effectiveness (Bearce & Tirone, 2010). Furthermore, the two objectives—geopolitical strategy and economic growth—are not necessarily mutually exclusive. Cold War donors may have wanted not only to strengthen political alliances with allies but also help them succeed economically. The strategic nature of aid to countries like South Korea and Taiwan during the Cold War did not blunt its effectiveness at improving economic and human development (Dreher, Klasen, Vreeland, & Werker, 2013). Thus, while liberal internationalists may disagree with the political motives behind realism and geopolitics, this approach is nonetheless a legitimate and established rationale for interstate relations.

While economists aimed to select countries based on quantitative measures of governance, economic policies, and human capital, some in the State Department wanted to channel funds to key allies. A senior Treasury official said, “State’s fear was that certain

countries they like for diplomatic reasons weren't going to get money and a bunch of others that we provide assistance to for national security reasons weren't going to pass the indicators. Some State officials worried that this cute little thing for development assistance was going to infringe on their ability to give money for political reasons" (Interviewee 1E, personal communication, June 28, 2015). Sometimes these foreign policy rationales were expressed in more parochial terms. A foreign policy community member was often aligned with a particular country or region rather than the indicators. A senior State Department official remarked, "The state regional bureaus had economists but they care a lot about the countries, they get clientitis. They care deeply about their offices and would always fall on the sword... They were advocates of their particular countries, not necessarily some universal logic" (Interviewee 1D, personal communication, June 22, 2015).

Department of Treasury staff, NSC International Economic Affairs members, and Deputies and Principals Committee members aligned with the White House's aid selectivity agenda—the economic policy community—were committed to their universal logic of growth *via* good governance. They disagreed with their foreign policy colleagues' geopolitical views. These conflicting rationalities set up "a policy dissonance at conception" (Interviewee 1H, personal communication, July 19, 2015). A tension existed in the MCA interagency process that was a microcosm of a broader rift between neoclassical economic and realist geopolitical rationalities. For example, Blackwill and Harris (2016b) stated that "[t]he neoclassical economic orthodoxy survived the Cold War, as did the resulting divide between economists and foreign policy thinkers" (Blackwill & Harris, 2016b: 107). Some of this tension was calmed by White House officials' promise that the MCA was going to supplement funding and not affect existing ODA levels. A senior Treasury official said, "The White House made the case that this was

about additional money, existing money would not be affected and it [the selection criteria] would only apply to the new MCA funds” (Interviewee 1E, personal communication, June 28, 2015). This calmed concerns somewhat but the US foreign assistance budget is notorious for being “a relatively tight budget with little wiggle room” (Interviewee 1G, personal communication, July 14, 2015). Tensions remained throughout the MCA creation process largely because the two communities’ rationalities were irreconcilable.

This raises the question of how individual donor countries like the United States were able to move their bureaucracies towards greater aid selectivity. In other words, how did the US executive branch change path-dependent patterns of distributing aid rooted in geopolitics to those rooted in “purer” development concerns? How did economic expertise wrest development assistance decision-making away from America’s foreign policy community to influence the distribution of the largest increase in bilateral development in forty years?

The economic policy community advocating for selectivity overcame the foreign policy community’s resistance through a strategy of cultivating a “trust in numbers” over foreign policy officials. Economic policy community members enrolled likeminded White House officials and Congressional representatives in their selectivity efforts. As part of this strategy, the economic policy community cast the foreign policy community as political and subjective while presenting themselves as apolitical and objective. The Treasury Department’s reputation as a center of objective, technical competence within the federal government aided the community’s strategy.

**“Breaking” the Foreign Policy Community.** Bush’s choice of Edson, the primary NSC official on international economic policy matters, to shepherd a foreign policy initiative like the MCA signaled the rising clout of economic expertise within the White House. *The New York Times* reported, “President-elect George W. Bush plans a major change in his National Security

Council, adding economics experts to involve it far more in the economic changes that have caused upheaval around the world” (Sanger, 2001: 1). Edson enlisted experts and leaders from the Cabinet departments and executive agencies such as State and Treasury as part of the interagency process and served as a liaison between Bush and those departmental officials.

Edson had informally reached out to and involved State and Treasury officials such as Alan Larson and Steven Radelet. There was no similar regard shown to USAID officials. In fact, since the agency was not a statutory NSC member, its participation was not guaranteed.

According to Reverend David Beckmann, President of Bread for the World, who was involved in the MCA process as a civil society advisor: “When the conversation started after the President’s MCA speeches, AID was not in the working group. They had to work like the dickens even to get invited to the meetings” (The Millennium Challenge Account, 2003: 78). Ultimately USAID was included in the interagency process but only because the actors involved, Patrick Cronin and John Simon, were political appointees and not career USAID bureaucrats. As one NSC official put it, “The NSC wanted to exclude USAID from the discussion but they accepted Cronin and John Simon. They were external to AID, not typical ‘AIDers’. They were viewed as acceptable and came from other sectors” (Interviewee 1B, personal communication, June 11, 2015).

Despite USAID’s diluted participation, this did not prevent the foreign policy community from raising realist or geopolitical arguments in the Interagency Group or putting up opposition to the use of universally applicable indicators for selection. Nor did it ensure that Congress would approve the Administration’s request for a substantial increase in foreign assistance funding. To make progress on these fronts, the economic policy community had to develop a strategy of “trust in numbers” (Porter, 1996) —strengthening relatively weak bureaucratic elites with quantitative methods to gain autonomy from mistrustful legislatures—through an alliance

with congressional representatives who favored selecting countries based on indicator performance.

The Bush administration thought it wise to shift decision-making, and thus accountability, away from foreign policy and aid bureaucrats. Another NSC official who was part of the interagency process said “[t]he president decided on a whole series of measures that would change the aid debate and hopefully make it easier to sell to a Republican congress and more conservative think tanks like the Heritage Foundation and American Enterprise Institute” (Interviewee 1M, personal communication, August 28, 2015). One of these measures was the indicator selection system, which the administration believed “would allow the executive branch some freedom and authority to do some things and they can say back that this is not just a Secretary of State slush fund. One way [Congress] can judge us is by these indicators” (Interviewee 1F, personal communication, July 2, 2015).

With help from the CGD, the MCA interagency Group built bridges with Republican legislators such as Senator Richard Lugar of Indiana and the US Senate Committee on Foreign Relations’ and House Committee on International Relations’ staff to sell the indicator selection system. The CGD strategically disseminated its proposal for a metrics-based country selection system to a “select group of congressional staffers who have responsibility for issues relating to U.S. development, foreign assistance, and international economic policies” at a luncheon on June 10<sup>th</sup> at the Capitol building (Weiner and Lucas, 2002). The economic policy community used this and other network-building efforts across Congress to triangulate against the foreign policy community and neutralize any opposition to an indicator-based selection system. While doing so, they cast the arguments of the foreign policy community as “subjective”, “political”, and rooted in personal interests while their arguments were cast as “objective”, “apolitical”, and free

of subjective discretion. They argued that the foreign policy community was too biased in their funding allocation decisions. An NSC official remarked, “Congress wanted to get away from a system where the Secretary of State has an itch it wants to scratch to give away money and Congress has to appropriate the funds” (Interviewee 1F, personal communication, July 2, 2015). Another official said, “A few billion more for USAID was not on the table. It would not have gone over well with Republicans” (Interviewee 1G, personal communication, July 14, 2015). The economic policy community capitalized on the fact that when it came to foreign aid “[t]he Hill wanted numbers demonstrating success” (Interviewee 1H, personal communication, July 19, 2015).

**The Rise of the Treasury Department in US Foreign Policy.** Treasury’s reputation was central to the economic policy community’s enrollment of allies. The Clinton Administration’s National Economic Council (NEC) set the stage for Condoleezza Rice’s recommendation that President Bush include the Treasury Secretary as a statutory member of the NSC. Established in 1993 to coordinate and monitor the federal government’s economic policymaking, the NEC was successful in “making certain that discussions of domestic issues included consideration of economic elements” (Seidman, 1996). In a memo to President Clinton, Robert Rubin, the NEC’s first director, put the NEC’s success in this way:

The early part of the NEC’s first year was spent establishing ourselves institutionally...to provide proper weight to economic consideration in areas that traditionally have underweighted economic impacts...better process should result in all agencies supporting whatever is agreed upon and should result in proper weighting of economic considerations in what have traditionally been thought of as non-economic areas (Rubin, 1993: 1).

The field of foreign assistance did not escape this rise of economists and economic rationales in US federal policymaking.



Once Treasury had a seat at the NSC interagency table, its reputation as the federal government's center for excellence on matters of international economics became an asset for involvement in an aid initiative whose objective was cast in terms of raising global economic growth. The economic policy community's appeals to the objectivity and neutrality of economic goals such as growth helped elevate Treasury over other actors in the MCA interagency process. One senior Treasury official recounted, "Treasury had a different kind of knowledge [than USAID], macroeconomic knowledge, development knowledge with the multilateral development banks. It was a policymaking body; a think tank" (Interviewee 1H, personal communication, July 19, 2015). The State Department and USAID also had international economics teams. Nonetheless, the Interagency Group perceived Treasury to possess the level and type of economic expertise necessary for MCA's creation. This included familiarity and comfort working with international financial institutions' economic researchers and quantitative data. A senior congressional staffer involved in MCA's creation put it this way: "Treasury had people who could look at numbers. It had PhD economists...they could hit the ground running, traditionally worked with the World Bank, so had those relationships and familiarity. A USAID ag[ricultural] economist who knows a lot about rice is not going to tell you about indicators" (Interviewee 1G, personal communication, July 14, 2015).

The interagency group decided that Treasury, given its human resources and brand of economic expertise, should lead the performance indicator-working group responsible for designing the country selection system. But stronger technical competency in quantitative work was not the sole reason. An equally or more important facet of Treasury's brand of economic expertise was its perceived objectivity and neutrality. Economists' familiarity with quantitative empirics has given them a widespread reputation as being scientific and therefore neutral and

apolitical. This characterization influenced the Bush White House's efforts to place trust in an indicator-based selection system rather than foreign policy bureaucrats. Funding allocation decisions based on governance had to be perceived as being rooted more in quantitative rules than subjective discretion. After all, John Taylor, the Treasury official in charge of the MCA indicator group, authored the famous "Taylor Rule".<sup>16</sup> Taylor strongly preferred a rule-based approach to policymaking: "If there is anything about which modern macroeconomics is clear however—and on which there is substantial consensus—it is that policy rules have major advantages over discretion in improving economic performance" (Taylor, 1993: 197). As one senior Treasury official put it, "Treasury would have been more bloodless about the indicators. State would have tried to game indicators and the selection process so that the countries they would want to give money to, the US's friends, would get selected" (Interviewee 1H, personal communication, July 19, 2015). In the end, neoclassical economic rationales and quantitative policy devices were successfully mobilized in the name and spirit of objectivity: "The policy people, mostly Edson, were busy fighting hard to keep the foreign policy folks out...the NSC econ[omics] leaders wanted to break the foreign policy approach of State and USAID. Sticking to numbers meant no subjectivity" (Interviewee 1B, personal communication, June 11, 2015).

As the indicator group's lead agency, Treasury became responsible for vetting indicators for the MCA country selection system. This work took place in Treasury's Office of the Deputy Assistant Secretary for International Debt, Development, and Quantitative Policy under the guidance of the Office of the Undersecretary for International Affairs. State Department and

---

<sup>16</sup> The Taylor Rule recommends that a central bank raise the nominal interest rate by a larger percentage than any percent increase in inflation. For example, a one percent increase in inflation should prompt a central bank to increase the nominal interest rate by more than one percentage point. The rule's goal is to foster macroeconomic price stability and full employment (Davig & Leeper, 2007).

USAID staffers played a supporting role on the Working Group by searching for indicators that Treasury could vet (Interviewee 1F, personal communication, July 2, 2015). The Debt, Development, and Quantitative Policy economists (“Treasury economists”) became obligatory passage points for any effort to operationalize the principle of selectivity in foreign assistance (Callon, 1984). As leaders of the indicator Working Group, they positioned themselves between the White House and the entire set of actors with a stake in selectivity. If the economic and development policy communities and Congress wanted to select countries in an “objective” way their efforts had to pass through Treasury economists’ methods.

### **Managing Dissent for Policy Expediency**

The economic policy community prevailed in enshrining indicator-based selectivity as a strategy but its members did not necessarily agree on which indicators were suitable for the task of ranking and selecting countries. Given the variety of indicators available, choosing the appropriate ones presented a “selection dilemma”. There was debate in the Interagency Group over which indicators to include the selection system, particularly over whether the World Bank’s WGI should be used to rate countries’ governance

**The MCA Selection System.** On November 25, 2002 the White House released a fact sheet outlining the key characteristics of the Bush Administration’s MCA initiative including country qualification criteria (White House, 2003). The quantitative qualification criteria, as determined by the Working Group and approved by President Bush, included per-capita income and indicators to judge countries’ performance. MCA qualification criteria also allowed the MCC Board—comprised of the Secretary of State, Secretary of the Treasury, US Trade Representative, MCC’s CEO, USAID Administrator, and four rotating civil society members—to exercise discretion in considering “other material information”, both quantitative and

qualitative, in making its final eligibility determination. Use of such material was typically reserved for countries that passed the performance indicators and qualified for funding (only a subset of qualifying countries are granted funding). Finally, the fact sheet mentioned that countries statutorily prohibited from receiving assistance under the Foreign Assistance Act were not eligible for MCA funds (White House, 2003).

The Bush Administration intended for the MCA to make substantial new financial resources available to low-income countries. Thus, it was necessary to restrict candidacy to relatively lower income countries in some way. The interagency group decided to use World Bank gross national income (GNI) per-capita data and guidelines to determine MCA candidacy. In the MCA's first year of operation, fiscal year 2004, countries that were allowed to borrow from the World Bank's International Development Association (IDA) and had per capita incomes below \$1,435, the IDA cutoff, would be considered candidates. In fiscal year 2005, all countries regardless of IDA eligibility with incomes below \$1,435 would be considered. In fiscal years 2006 and beyond, all countries with incomes up to \$2,975, the World Bank cutoff for lower middle-income countries at the time, became candidates. Beginning in 2006, separate competitions would be run for countries with incomes below \$1,435 (the "low-income" group) and those with incomes between \$1,435 and \$2,975 (the "lower middle-income" group) (White House, 2003).<sup>17</sup> Countries that met these income criteria and were not barred by the Foreign Assistance Act from receiving U.S. assistance were called "MCA candidate countries".

Once candidacy was determined, countries would have to compete with their income category peers on a set of performance criteria. The interagency group settled on 16 indicators (see Table 4.1). To qualify as a "better performer" a country would have to score above its

---

<sup>17</sup> The income cutoffs changed for every subsequent year since 2006 for both groups.

peers' median score on half of the indicators in each of three policy areas: Ruling Justly, Investing in People, and Economic Freedom.<sup>18</sup> These categories corresponded to what the Council of Economic Advisors concluded, based on neoclassical economic theory and empirics, were the primary determinants of growth (governance, human capital, and macroeconomic policies, respectively). These categories, or “buckets” as Interagency Group members called them, bore similarities to categories used in multilateral banks' institutional and policy assessments. The World Bank uses a Country Policy and Institutional Assessment (CPIA) to determine allocations in its concessional lending and grant-making program for relatively lower income countries (i.e., IDA countries).<sup>19</sup> Regardless of its performance on other criteria, a country also had to score above its peers' median score on the WGI Control of Corruption indicator (White House, 2002). This requirement was referred to colloquially as the “corruption hard hurdle”. Candidate countries passing the performance indicator criteria “qualified” for MCA funding.

---

<sup>18</sup> There was one exception to countries competing against its peers. Countries were not required to score better than the group's median score on the inflation indicator but instead had to fall below a threshold, usually 15 percent.

<sup>19</sup> This is an internal assessment conducted by World Bank staff using 20 criteria in four categories: Economic Management, Structural Policies, Policies for Social Inclusion/Equity, and Public Sector Management and Institutions. The MCA's Ruling Justly category was like the CPIA's Public Sector Management and Institutions category. Investing in People bore some resemblance to the CPIA's Policies for Social Inclusion/Equity category. The MCA's Economic Freedom category was similar to the CPIA's Economic Management and Structural Policies Categories.

Table 4.1

*Original MCA Qualification Criteria*

Indicator	Source
<b>I. Ruling Justly</b>	
1. Control of Corruption	World Bank Institute
2. Rule of Law	World Bank Institute
3. Voice and Accountability	World Bank Institute
4. Government Effectiveness	World Bank Institute
5. Civil Liberties	Freedom House
6. Political Rights	Freedom House
<b>II. Investing in People</b>	
7. Immunization Rate: DPT and Measles	WHO/World Bank
8. Primary Education Completion Rate	World Bank
9. Public Primary Education Spending/GDP	World Bank
10. Public Expenditure on Health/GDP	World Bank
<b>III. Economic Freedom</b>	
11. Country Credit Rating	Institutional Investor
12. Inflation	IMF
13. Regulatory Quality	World Bank Institute
14. Budget Deficit/GDP	IMF/World Bank
15. Trade Policy	Heritage Foundation
16. Days to Start a Business	World Bank

*Source:* Radelet (2003), *Challenging Foreign Aid: A Policymaker's Guide to the Millennium Challenge Account*, Washington, DC: Center for Global Development.

The Working Group began meeting at least once a week in the White House Situation Room starting in the spring and then throughout the summer of 2002 to narrow down the set of indicators used to rank and choose countries (Interviewee 1J, personal communication, August 5, 2015). The Working Group was responsible for vetting indicators and coming up with a shortlist for the wider Interagency Group to consider and approve (see **Figure 4.5**). At least 70 possible indicators were originally considered, including over 25 governance indicators. The foremost desired indicator characteristics included: 1) transparency and independence, 2) positive relationship to economic growth, 3) high country coverage, and 4) measurement validity (Treasury, n.d.). Several of these criteria were outlined in a September Center for Global

Development report, *On Eligibility Criteria for the Millennium Challenge Account* (Birdsall, Levine, Lucas, & Shah, 2002).

The Interagency Group considered an indicator transparent and independent if a non-US government, third-party source produced it (if the indicator and perhaps its underlying data were publicly available, that was considered a bonus). The stronger an indicator's positive correlation was with economic growth the more favorable it appeared to the group. Likewise, the greater an indicator's country coverage, or whether it had data measuring a country's performance, the better. The group also considered it important that an indicator represent the spirit of what its respective bucket intended to measure, i.e. an indicator should be an accurate representation of governance, investment in human capital, or free market economic policies.

**Figure 4.5.** Indicator Work Plan Circulated to the MCA Interagency Group Members

<b>Draft Work Plan</b>	
<b>Performance Indicators Working Group</b>	
<b>April 5, 2002</b>	
<u>Meeting Date</u>	
April 5, 2002	Discuss primary sources of indicator data. Discuss specific criteria for indicator selection. Discuss if indicators are for country selection only or will also be related to allocation.
April 11, 2002	Discuss/select broad set of indicators for consideration.
April 18, 2002	Select initial indicators.
April 25, 2002	Discussion of how to combine and interpret indicators
May 2, 2002	Discuss initial results of rankings and correlations for selected indicators. (Rankings will be made available before meeting; agencies should come to meeting with criticisms/concerns.)  Determine methodology for scoring/selection.
May 9, 2002	Presentation of revised rankings. Develop next steps.

*Source:* Performance Indicator Working Group (2002a). *Draft Work Plan. April 5, 2002.* Washington, DC: US Department of the Treasury.

**Governance Indicators as “Plywood.”** When choosing indicators for the three specific buckets, there was little disagreement over proposed indicators for the Investing in People and Economic Freedom categories. Selecting indicators for the Ruling Justly category proved contentious, however. Academic economists in the economic policy community were skeptical of the WGI’s aggregation methodology and thus its measurement validity as well. The WGI, however, was more suitable for operationalizing the MCA’s global indicator-based competition for funds. When it came to the WGI’s use in the selection system, policy expediency trumped methodological concerns.

There was broad agreement that there had to be a mix of input and outcome indicators in the Investing in People category and the Interagency Group felt the proposed indicators on public expenditures, immunization rates, and education completion rates met the relevant criteria. The Economic Freedom indicators were reminiscent of the earlier structural adjustment era. Five of the indicators were focused on what might be traditionally considered macroeconomic criteria. They were either the same indicators or close proxies examined by structural adjustment programs that sought to stabilize countries’ macroeconomic environments, liberalize their markets for finance, goods, and services, and deregulate industry and production. One indicator, “Days to Start a Business”, was novel and focused on the microeconomic environment. There was general agreement on these indicators even if some interagency group members did harbor concerns about these indicators’ “free market bias”. As one Interagency Group member from Treasury put it, “NSC International Economics was very committed to the ‘rising tide lifts all boats’ approach to growth...there was a commitment to the idea that growth was a good thing and should be encouraged and that the free market approach was the best way to do that” (Interviewee 1J, personal communication, August 5, 2015).



The primary issue for the proposed Ruling Justly indicators was whether the indicators that made up the majority of proposed criteria in this bucket—the WGI—possessed sufficient measurement validity. After all, it was Douglass North who said, “We cannot see, feel, touch, or even measure institutions; they are constructs of the human mind” (North, 1990: 107). These essentially qualitative mental constructs had to be translated into quantitative measures like the WGI using database generation and modern statistical models. This was not without its pitfalls: “It is useful to remember that governance is essentially a qualitative phenomenon, the quantification of which would always be subject to considerable empirical limitations” (Malik, 2002: 3). John Taylor, who was directing the Working Group on behalf of Treasury, had reservations about the WGI’s empirical limitations.

The WGI’s six indicators are each a composite index of governance. When the WGI was first compiled in 1996 it included eleven sources. By 2011 the number of sources had grown to thirty one (Kaufmann, Kraay, & Mastruzzi, 2011). Kaufmann and Kraay constructed each of the six aggregate WGI indicators by averaging together data from underlying sources they believed corresponded to the respective composite indicator’s governance category. This was (and still is) done in three main steps. First, individual questions from the underlying producer data source are assigned to each of the relevant aggregate indicators; for example, questions on bribery are assigned to the Control of Corruption composite index. Next, questions from the individual data sources are rescaled from 0 to 1 with higher values corresponding to more preferable outcomes; for example, a score of 4 on a question ranging from 1 to 5 is rescaled as 0.75. This rescaling does not make individual sources comparable to each other, however. This is accomplished through the last step, the construction of a weighted average of the individual source indicators with a statistical methodology known as an Unobserved Components Model. The resulting

governance indicators represent a weighted average of each source's data, with Kaufmann and Kraay assigning greater weight to sources that tend to be more strongly correlated with each other. The final composite indicators' values ranged from -2.5 to 2.5, with higher values corresponding to more preferable outcomes (Kaufmann, Kraay, & Zoido, 1999a).

The Unobserved Components Model is essentially an aggregation method. The WGI brings together governance indicators that differ in terms of the governance concept measured, nature of respondents providing information, sample of countries they cover, and units of measurement (Kaufmann, Kraay, & Zoido, 1999b). Kaufmann and Kraay's primary motivation for aggregating these diverse indicators was to provide a more precise measure of governance than individual indicators could to facilitate formal hypothesis testing of cross-country differences in governance (Kaufmann et al., 1999a). Their argument is that individual governance indicators are imperfect signals of actual existing governance. Given the diversity among indicators and variety of governance definitions, the measurement error of each individual indicator is quite high (e.g., can a single survey of households' perceptions of bribe payments reflect a country's corruption?). Combining indicators, however, can theoretically get us closer to the true nature of a country's governance, or in statistical laymen's terms, improve the signal relative to the noise:

Since "true" governance is difficult to observe and we can observe only imperfect indicators of it, how can we best extract a "signal" of unobserved governance from the observed data? Under this view, all individual indicators of corruption, for example, should be viewed as noisy or imperfect proxies for corruption. Aggregating these together can result in a more informative signal of corruption (Kaufmann et al., 2011: 238).

Treasury Undersecretary John Taylor took issue with the WGI's aggregation process. One of his associates recalled that "[h]e referred to the indicators as plywood—it looked like wood, but it wasn't really. It was a bunch of layers of measures laid on top of measures to look

like wood but wasn't nearly as strong as wood" (Interviewee 1B, personal communication, June 11, 2015). Also, Taylor reportedly said, "Is what you are measuring in Tanzania the same as in Kenya? They say it is Tanzania, but are we sure that bureaucratic efficiency in Tanzania is the same as bureaucratic efficiency in Kenya, and should we be comparing those two? Even more so between say Vietnam and Tanzania. Are these indicators capable of flattening and presenting the concept as the same across such varied environments?" (Interviewee 1B, personal communication, June 11, 2015). Taylor's main contention was that separate indicators and measures making up the index should not be aggregated because they used different surveys in different countries that lacked a standardized approach.

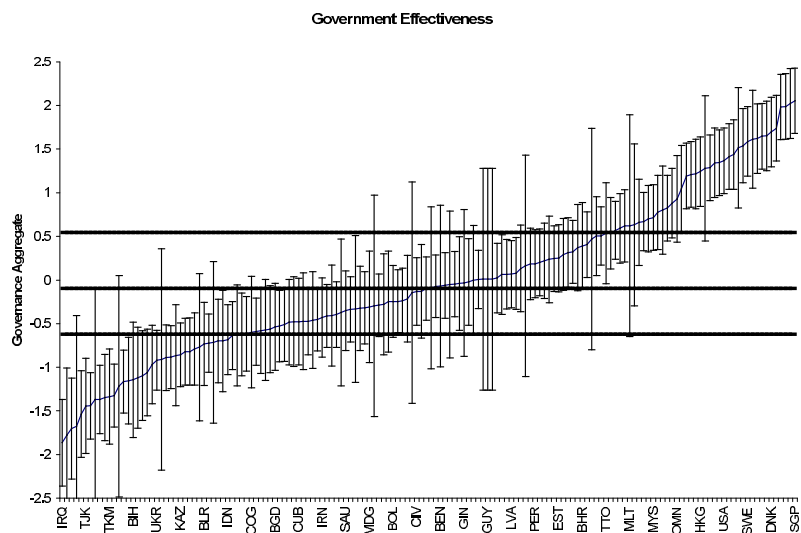
If the WGI's estimates raised questions about measurement validity, then this posed a challenge for selecting countries based on performance. How would the MCA architects know if a country was actually performing better than its peers on the WGI? On this issue, Taylor was not alone. Kaufmann and Kraay did not shy away from their composite indicators' shortcomings. They candidly concluded that "[d]espite several optimistic assumptions, we find that governance is not very precisely measured using these aggregate indicators. In particular, although it is possible to identify statistically significant differences between countries at opposite ends of the distribution of governance, it is much more difficult to discriminate among the majority of countries with any degree of confidence" (Kaufmann et al., 1999a: 27). Aggregation did little to improve measurement errors, which remained high. Consequently, confidence intervals around the composite indicators' point estimates were very large. This made comparison of governance across countries practically meaningless. When the distribution of WGI's governance estimates was split into quartiles, large confidence intervals made it

difficult to say with any statistical certainty where in the distribution a particular country belonged (see **Figure 4.6**):

For a small group of countries at each end of the distribution of governance, we can conclude with a great deal of confidence that these countries are in fact in the top and bottom quartiles. However, for the middle quartiles the situation is much less clear, as very few countries' 90% confidence intervals lie entirely within a given quartile... In particular, for the 'typical' country around the middle of the distribution of governance, governance is not significantly different from nearly half of all other countries in the world (Kaufmann et al., 1999a: 16–18).

In other words, the WGI could, for example, confidently conclude that governance was “better” in Scandinavia than Central Africa, but precious else about differences in the rest of the world.

**Figure 4.6.** Plot of Countries' WGI Government Effectiveness Point Estimates with Confidence Intervals and Quartile Distributions



*Source:* Kaufmann, D., Kraay, A., & Zoido, P. (1999a). Aggregating Governance Indicators (SSRN Scholarly Paper No. ID 188548). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=188548>

*Notes:* The dotted line represents the country's point estimate and the vertical bars represent its confidence interval. The horizontal lines define cutoffs for the quartiles. Some countries' confidence intervals were so large (e.g., Guyana) they could fall in the same quartile as the country with the worst point estimate, Iraq, or best point estimate, Singapore.

Kaufmann and Kraay were concerned enough about measurement error that they wrote a discussion paper specifically for the MCA interagency group. In the paper they generally concurred that it was worthwhile to measure governance and choose countries based on quantitative criteria, but also sounded the alarm that the WGI should not be used to run “horseraces” between countries to determine MCA eligibility: “It is difficult to assign [many countries] with a high degree of confidence to a definitive performance category according to their estimated level of governance” (Kaufmann & Kraay, 2002: 4). As a result, there is a “non-trivial probability” that countries could be mistakenly disqualified; the authors recommended the use of supplemental in-depth, country governance diagnostic surveys to make qualification decisions (Kaufmann & Kraay, 2002).

While Taylor, Kaufmann, and Kraay were all interested in influencing policy, they wished to do so without betraying the limitations of social scientific methods. They were in favor of applying these methods to policy so long as doing so respected the boundaries of what the evidence and tools reasonably allowed. Their first commitment was to ontological objectivity, or getting as close to the ultimate structure of reality as possible (whether such a governance reality does or can exist is another matter). Senior economic policymaking officials in the NSC were not interested in methodological debates or ontological objectivity. They were interested in whether the final outputs of social science methods could further their goals, in this case, providing a set of indicators that united the Administration’s good governance and growth agendas while selecting countries for funding.

A divide began to emerge in the MCA interagency group between academic research economists on one side and the economic policymakers on the other. In early May 2002 Taylor instructed some of his Treasury staff members on the indicator-working group to come up with

Treasury’s own governance indicator as an alternative to the WGI. Taylor wanted to substitute the WGI with an unambiguous set of specific survey questions that captured MCA’s governance priorities while providing greater transparency (Treasury 2002a, 2002b). Ten questions were extracted from three out of twenty four possible sources: the World Economic Forum’s Executive Opinion Survey, Columbia University’s State Capacity Project, and World Bank’s Business Environment Survey (see **Figure 4.7**).

**Figure 4.7.** Individual Components of Treasury's Proposed Composite Governance Index

**Indicator Construction**

The Treasury indicator aggregates answers to ten survey questions – two from each of the following areas: public expenditure management, rule of law, property rights, corruption, and leadership.

	Type of Question				
	Public Expenditure Management	Rule of Law	Property Rights	Corruption	Leadership
WB, World Economic Forum	1	1	1	1	1
WB, World Business Environment Survey	1	0	1	0	0
Columbia University, State Capacity Project	0	1	0	1	1
<b>Total # Questions</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>

**Treasury Indicator Questions**

1. Do unfair or corrupt activities of other firms impose costs on your firm? - 2001 WEF (CR)
2. Rate the severity of corruption within the state - 1999 SCS (CR)
3. Public trust in the honesty of politicians - 2001 WEF (LD)
4. Admin/ technical skills of the country's primary political decision-makers (e.g. chief exec, cabinet) - 1999 SCS (LD)
5. The composition of govt spending in country is wasteful/ useful - 2001 WEF (PE)
6. How would you rate the efficiency of central/local govt in delivering services now vs. 3 yrs ago - WBES (PE)
7. Protection of financial assets and wealth - 2001 WEF (PR)
8. Confidence that the legal system will uphold by contract and property rights in business disputes" - now vs 3 yrs ago -- WBES (PR)
9. Judicial independence - 2001 WEF (RL)
10. State's adherence to the rule of law - 1999 SCS (RL)

*Source:* Treasury (2002b). *Treasury Governance Indicator*. Washington, DC: US Department of the Treasury.

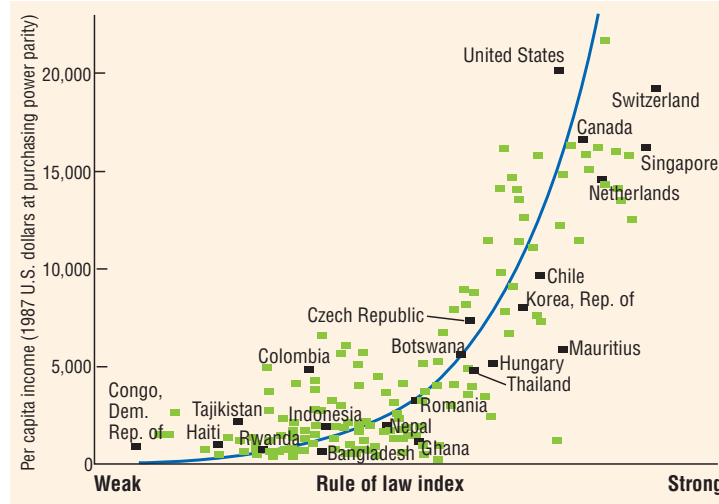
The Center for Global Development, which was influential early in the MCA process in setting the standards for “optimal” eligibility criteria, had a different understanding of transparency than Taylor did, however. The Center for Global Development recommended that any eligibility criteria “be accessible to non-governmental organizations and researchers in the

U.S. and potential recipient countries. This is the only means to achieve the transparency and clarity that will distinguish the MCA” (Birdsall, Levine, et al., 2002: 6). According to CGD, indicators were transparent if they were *public* and *accessible*. Taylor, meanwhile, believed indicators were transparent if they were *accurate* and *valid*. NSC economic policymakers’ found common cause with CGD’s conception of transparency. The Treasury index produced by a small group of government economists was riskier because it could be cast as private and provincial. Unlike the World Bank’s WGI, the Treasury index could more easily be branded a “US government product” and associated with narrow US economic or foreign policy interests (Interviewee 1B, personal communication, June 11, 2015). NSC officials needed to project objectivity as strongly as possible to overcome opposition from the foreign policy community. Indicators perceived as public, accessible, and independent could be contrasted as a more righteous alternative to the foreign policy community’s parochial geopolitical concerns and more easily gain the support of Congress and development practitioners.

The Treasury index also had not been “externally validated” in the same way that WGI had been. The CGD report advised that “[d]ata sources that have been used extensively in research study, and indicators that have been correlated to outcomes of interest, are far more desirable than brand new data sources that have had no external validation” (Birdsall, Levine, et al., 2002: 6). The WGI met this criterion. Kaufmann and Kraay had studied their indicators’ relationship to growth in World Bank Policy Research Papers. They later disseminated these findings in an International Monetary Fund magazine, *Finance & Development*, through stylized facts that clearly depicted their indicators’ relationship to GDP per capita (see **Figure 4.8**). Moreover, Treasury economists conducted econometric work similar to the World Bank’s to determine indicators’ relationship with growth. (Interviewee 1J, personal communication,

August 5, 2015) (see **Figure 4.9**). The WGI had a strong positive relationship with countries' per capita income. This was an important part of the vetting process because indicators lacking this positive relationship with economic growth were not considered for inclusion.

**Figure 4.8.** Stylized Fact Representing Positive Relationship between WGI's Rule of Law Indicator and Per Capita Income



Source: Kaufmann, D., Kraay, A., & Zoido-Lobaton, P. (2000). Governance matters: From measurement to action. *Finance and Development*, 37(2), 10.

**Figure 4.9.** Example of Analysis Conducted by Treasury Economists in the Indicator Working Group

<b>Governance Correlations- IDA &amp; IDA Blend Countries</b>	CIVILLIB	POLRIGHTS	KKVOICE01	KKGOV01	KKRULLAW01	KKCOR01	GDP1	GDP2	POVERTY
CIVILLIB	1.00	0.88	-0.84	-0.37	-0.47	-0.26	-0.23	-0.07	-0.06
POLRIGHTS	0.88	1.00	-0.85	-0.26	-0.47	-0.25	-0.21	0.04	-0.12
KKVOICE01	-0.84	-0.85	1.00	0.47	0.62	0.25	0.20	0.12	0.09
KKGOV01	-0.37	-0.26	0.47	1.00	0.62	0.46	0.41	0.19	-0.03
KKRULLAW01	-0.47	-0.47	0.62	0.62	1.00	0.48	0.27	0.15	-0.07
KKCOR01	-0.26	-0.25	0.25	0.46	0.48	1.00	0.30	0.09	0.07
GDP1	-0.23	-0.21	0.20	0.41	0.27	0.30	1.00	0.11	0.24
GDP2	-0.07	0.04	0.12	0.19	0.15	0.09	0.11	1.00	-0.28
POVERTY	-0.06	-0.12	0.09	-0.03	-0.07	0.07	0.24	-0.28	1.00

Source: Council of Economic Advisors (2015). Kroszner, Randall (Randy) – Subject Files in *Trade: Millennium Challenge Account*. Dallas, TX: George W. Bush Presidential Library and Museum.

Notes: Variables starting with “KK” are WGI indicators. The variable GDP1 refers to countries' GDP per capita between 1995 and 1999. GDP2 refers to countries' GDP per capita between 1990 and 1995.



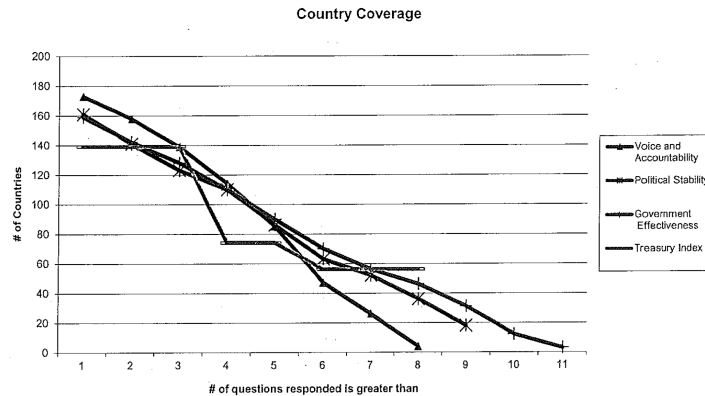
Finally, there was the issue of country coverage. The WGI covers “virtually all countries in the world” (Kaufmann, Kraay, & Zoido-Lobaton, 2000). In 1999 the WGI covered 155 countries and by 2009 it covered 211, which permitted comparisons, albeit imprecise, across a much larger set of countries than was possible using any single indicator (Kaufmann et al., 2011). Country coverage was instrumental to operationalizing a global competition to satisfy performance-oriented conservatives. According to an NSC official, “There were tensions over data quality and coverage. Coverage won out because the selection system required wide country coverage to work” (Interviewee 1B, personal communication, June 11, 2015). Choosing countries for funding required comparing countries’ scores in a standardized, universal way. Countries whose governance and other characteristics could not be seen could not be properly evaluated. This form of global aid management requires the state—in this case the US federal government—to “see” the globe (Scott, 1998). This applies not only to state practices within but also across borders (Broome & Seabrooke, 2012). The WGI, because of its successful aggregation process, was best positioned to see and evaluate the largest number of countries (see **Figure 4.10**).

**Figure 4.10.** Country Coverage Comparisons Between WGI and Treasury's Proposed Indicator

**Data Coverage by Index for Governance Measures (all)**

Countries Per Question	Voice and Accountability		Political Stability		Government Effectiveness		Treasury Index	
	Resp.	Min	Resp.	Min	Resp.	Min	Resp.	Min
0	2	175	14	175	16	175	62	175
1	15	173	20	161	17	159	0	139
2	19	158	18	141	14	142	6	139
3	25	139	13	123	18	128	41	139
4	29	114	24	110	20	110	0	74
5	38	85	23	86	20	90	32	74
6	21	47	11	63	14	70	0	56
7	22	26	16	52	10	56	10	56
8	4	4	18	36	15	46	15	56
9			18	18	19	31	0	
10					9	12	41	
11					3	3	0	

Resp. is the number of countries that have exactly that many questions covered.  
Min. is the number of countries that have responses to at least that many questions.



Source: Treasury (2002b). *Treasury Governance Indicator*. Washington, DC: US Department of the Treasury.

For these reasons, “The majority of the working group did not favor the Treasury Governance indicator as an option to replace the Kaufmann-Kraay indicators” (Performance Indicator Working Group, 2002b: 2). Despite Kaufmann and Kraay’s admission that the WGI did not precisely measure governance and the composite indices were, at best, modest improvements over individual governance indicators, the WGI still presented the best option to instrumentalize governance indicators for aid allocation. Taylor ultimately consented to the WGI even if he still held reservations and the Interagency Group came to accept that a bad governance measure was better than no measure at all. Treasury economists, with help from the

CGD report and guidelines, successfully mediated between Taylor's commitments to ontological objectivity and the NSC economic policymakers' commitments to policy expediency.

### **The Power of Method: Statistical Consensus and Concealing Bias**

Contestation over which indicators to use for selection demonstrated that while the economic policy community bemoaned the foreign policy community's selection bias, in reality bias had shifted from the selection of countries to the selection of indicators and the biases these indicators possessed. The WGI's use had an effect on the nature of decision-making as well as the distribution of development finance. It collapsed incredibly complex country politics into a singular metric that ignored important factors for governance such as years since independence and institutional path dependencies. The WGI's inclusion in the MCA selection system meant that decision-making was shifted away from the Washington, DC foreign policy community and towards an assemblage of country risk analysts, survey instruments, and statistical models. In the end, the WGI's incorporation as a decision-making tool for aid allocation did not lead to rationality taming subjectivity but swapped the subjectivity of foreign policy experts for the subjectivity of a diverse and distributed network of mostly country investment risk analysts.

The indicators included another type of bias—income bias. Wealthier countries routinely performed better on the WGI than did relatively poorer ones. As the remainder of this chapter will show, countries that passed four or more WGI indicators and qualified for MCA assistance, were, on average, wealthier than those that did not. Because WGI scores barely changed over time, this resulted in a set of countries that became a “global underclass” when it came to eligibility for global development finance.

**The Changed Nature of Selection Bias.** The WGI aggregates the massive diversity of governance conditions around the world into a simplified and standardized common metric.

Commensuration transforms different qualities into a single quantity. It converts difference into magnitude (Espeland & Stevens, 1998). When many different governance data sources are aggregated into a single composite index, it makes it easier to talk about “more or less governance” rather than different kinds of governance (Espeland & Stevens, 1998). Tellingly, Kaufmann and Kraay refer to their successful commensuration as a “statistical consensus”, “Although different sources measure governance in very different units, statistical techniques are available that allow us to anchor each source in a common set of units” (Kaufmann et al., 2000: 11).

This statistical consensus foreclosed discussions around the context in which countries’ institutions took shape. Take the case of countries in the Central Asia region, for instance. The MCA’s reliance on the WGI resulted in a rejection of countries in this low-income region of the world. Central Asia consists of three low-income countries, Kyrgyzstan, Tajikistan, and Uzbekistan, and one country that transitioned from low-income to lower-middle income, Turkmenistan (Kazakhstan is part of Central Asia but was only considered in 2006 and 2007 after which it graduated from the lower middle-income country category and was no longer a candidate for MCA funding). Nine years of potential MCA qualification across Central Asian countries represented 36 possible instances of qualification. Failure to score above the median on the WGI’s four indices in the Ruling Justly category disqualified Central Asian countries exactly half the time (18 instances) (see Table 4.2). These countries nonetheless passed the other two categories of Investing in People and Economic Freedom the majority of the time. No other region’s countries exhibited this pattern so often. Kyrgyzstan, Tajikistan, and Uzbekistan were the most affected. Uzbekistan, however, was regularly statutorily prohibited from receiving funds because of the US Congress’ and State Department’s concerns of human rights violations.

Table 4.2

*Central Asian Countries Disqualified from MCA Consideration Due to WGI Scores*

Country	GNI / capita in 2008 (US \$)	Number of Years Failed to Qualify due to WGI	Years in Which Qualification Prevented due to WGI
Kyrgyzstan	590	6	2004–08, 2011
Tajikistan	460	5	2004, 2006, 2007, 2009, 2011
Uzbekistan	730	5	2004–08
Turkmenistan	3040	2	2007–08

*Source:* Based on author's calculations of MCA performance data. GNI per capita data taken from the World Bank's Worldwide Development Indicators.

Irrespective of whether Central Asian countries would generate a good “return” on US bilateral assistance, reliance on governance indicators like the WGI obviated questions and discussions that might take countries' or regions' unique history into account. The WGI neither accounts for when these and other countries began their own efforts at improving governance nor their specific institutional path dependencies. Former colonies in the Global South gained their independence from different empires and as a result inherited vastly different institutional legacies.

Central Asian countries are nestled in between Russia, China, and Iran. Every Central Asian country was a once part of the Soviet Union as a Soviet Socialist Republic until it achieved its independence in 1991. Many post-Soviet Union countries in Eastern Europe, the Caucasus, and Central Asia performed well in the Investing in People category because the Soviet Union invested heavily in basic literacy and health across its republics. The immediate post-Cold War era saw several former Soviet Union territories' administrations, including those in Central Asia, move quickly to adopt market-based economic reforms. These countries routinely scored above the median on MCA's macroeconomic indicators measuring inflation,

fiscal deficits, and trade openness. At the same time, the legacy of the Communist Party of the Soviet Union presented path dependencies for Central Asian countries' governance, institutions, and norms. Many Central Asian administrations were slow to adopt democratic reforms and embrace freedom of association, opposition, and the press. By the end of the 1990s, Central Asian countries' administrations had built a reputation for being authoritarian with high levels of corruption and uneven property rights' definition and enforcement. As a result, countries in the region received negative governance assessments and routinely fell below the median on all four Ruling Justly WGI indicators. This precluded them from MCA assistance.

A discussion about Central Asian countries' unique context might have been had by the Board to select eligible countries from the pool that qualified for assistance (i.e., passed the scorecard) but not for those that, for example, failed all four WGI indicators in the Ruling Justly category. The WGI's statistical consensus made it more difficult to consider important, nuanced contextual factors affecting countries' governance scores. In weighing funding decisions the foreign policy community might have considered, for example, that Central Asian countries were some of the last to have achieved independence among MCA candidate countries. By contrast, many West and Central African countries gained their independence from France in 1960. Most former Portuguese African colonies gained their independence in 1975. Factors such as years since independence and institutions inherited at independence influenced performance on the MCA indicators. Having an extra fifteen or thirty years of independence could have a material affect on countries' actual institutions or how institutions are perceived. Likewise, reform programs that grapple with a legacy of French or Portuguese colonial institutions versus Soviet communist institutions presented very different options and paths for countries to follow.

The indicators included another type of bias—income bias. Choosing indicators based on their positive relationship with per capita income proved problematic because it meant richer countries would have a higher probability of scoring better on performance indicators. The interagency group became caught between its commitment to indicators that had a positive relationship with growth, largely to appease the economic policy community, and rising concern over what they called income bias: “Countries with higher per capita income, can that have a bias towards a specific indicator? Answer is almost always ‘yes’ for most of these indicators...there was an income bias” (Interviewee 1F, personal communication, July 2, 2015). Income bias is a problem because “to the extent that income levels themselves are driving a countries [sic] rating on the indicator, it raises the possibility that countries are being (indirectly) penalized simply for being poor. In such cases, the MCC selection process would be biased...low income would cause poor performance on the indicators and lower likelihood of qualification” (MCC Development Policy, 2005: 36).

This potential built-in bias against poorer countries concerned many members of the interagency group, particularly those in the NSC Office of African Affairs. Sub-Saharan Africa, the region with some of the lowest per capita incomes in the world, would be very affected by any income bias. Jendayi Frazer, the NSC Office of African Affairs’ director, and some of her colleagues were concerned that something countries might not be able to control, such as geography, was related to performance in a way that unduly punished them (Interviewee 1B, personal communication, June 11, 2015). One NSC official remarked, “What if all your neighbors are poor or all your neighbors are rich? Do your neighbors affect your governance, and if you happen to be surrounded by poor countries, are you really better at governance than the indicators suggest? Similarly the opposite when surrounded by rich countries” (Interviewee 1B,

personal communication, June 11, 2015). The Brookings Institution shared the NSC Office of African Affairs' concerns and made them public in a 2003 policy brief titled *Making the Millennium Challenge Account Work for Africa* (Brainard & Driscoll, 2003).

The NSC Office of African Affairs began conducting its own regression analysis to see if Africa was discriminated against on the basis of income. On the basis of this analysis the NSC Office of African Affairs raised its concerns to National Security Advisor Condoleezza Rice in an official memo outlining the risks of income bias and its potential effect on Sub-Saharan African countries' selection (Interviewee 1B, personal communication, June 11, 2015). The Interagency Group ultimately decided to address the issue of income bias by creating separate competitions for low-income and lower middle-income countries. Countries' performance on the indicators would be assessed against their per-capita income peers.

The WGI's inclusion in the MCA selection system meant that households' and experts' subjective perceptions would inform how billions of dollars of aid funding would be distributed. There were Interagency Group members who took objection to the decision: "Jendayi Frazer has worked on the ground [in countries]. She was skeptical of how surveys of governance were done. She knew it was done by some guy sitting in a café filling out a survey" (Interviewee 1B, personal communication, June 11, 2015). Kaufmann and Kraay acknowledged these types of concerns but found their exclusive use of perceptions based data to measure governance completely acceptable and somewhat inevitable.

The WGI's overreliance on investment risk analysts is important because such experts are prone to bias.<sup>20</sup> The largest contributors of governance information in the WGI are what

---

<sup>20</sup> The unobserved components model assumes that measurement errors in individual governance indicators making up the composite indices are uncorrelated with each other. In reality, measurement errors are likely related because



Kaufmann and Kraay refer to as “commercial business information providers” (Kaufmann et al., 2011). In 2009, among the WGI’s 31 sources, the five investment risk analyst sources alone contributed one-third (34 percent) of all data points. This does not include a few surveys, such as the World Economic Forum’s and Institute for Management and Development’s competitiveness surveys, which pose similar questions to similar groups of experts. Meanwhile, the nine NGO sources account for only 20 percent of all data points. More importantly, the five investment risk analyst firms have the highest country coverage among the various sources (Kaufmann et al., 2011). This means that for many countries, usually the smallest and poorest, they are one of the few sources of governance data available.<sup>21</sup>

An analysis of risk experts at *Institutional Investor*, a country risk appraisal guide similar to those used in the WGI, showed that its analysts had an overly adverse view of less-developed countries in the late 1980s (Somerville & Taffler, 1995: 293).<sup>22</sup> Furthermore, analysts may be influenced by “sentiment” about particular countries rather than data on economic fundamentals. When this sentiment is widely shared, it becomes a herd effect and generates bias (Somerville & Taffler, 1995). Kaufmann and Kraay themselves acknowledged this potential for “halo effects” when one-off economic or political events or recent experiences with growth affect an analyst’s perception of governance (Kaufmann et al., 2011).

---

experts tend to follow each other’s opinions, are more pessimistic than optimistic, and unduly influenced by high profile, one-off events.

<sup>21</sup> The unobserved components model is “unbalanced”. A balanced model would mean that all comparisons across countries and time would rely on the same set of data sources. An unbalanced model, by contrast, includes a country in a given year for an aggregate index if *any* individual indicator has an observation for that country in the given year. This greatly expands the number of WGI’s observations across space and time but increases the risk that changes in governance are simply the result of changes in the underlying data.

<sup>22</sup> Incidentally, *Institutional Investor* was an indicator source in MCA’s Economic Freedom category

The WGI's inclusion in the MCA selection system meant that decision-making was shifted away from the Washington, DC foreign policy community and towards an assemblage of country risk analysts, survey instruments, and statistical models. Kaufmann and Kraay's quantification and aggregation processes lent the WGI an aura of objectivity. In the end, however, the WGI's incorporation as a decision-making tool for aid allocation swapped the subjectivity of foreign policy experts for the subjectivity of a diverse and distributed network of mostly country investment risk analysts. Consequently, a theoretical claim about the appropriate standards for evaluating governance accompanies this assemblage: what is good for private foreign finance is good for official development assistance.

### **The Power of Method: The Distribution of Development Funding**

**The Deserving and Undeserving Poor.** Indicators incorporated into the country selection system affected which countries qualified for MCA funding. Qualification was determined based on countries' performance on the Interagency Group's chosen indicators summarized on a scorecard. Qualification did not guarantee MCA eligibility, however. Every year there were more countries that qualified than were made eligible. The MCA Board of Directors determined final MCA eligibility among qualifying countries using data on indicator performance as well as supplemental information.

The Board could use its discretion to qualify countries that did not pass the scorecard. However, in only two cases in MCC's history did the Board deem a country eligible in its first year of qualification without it having passed the indicator criteria: Georgia and Mozambique, both in 2004. Between 2004 and 2015, ninety-three percent of countries selected by the Board met the indicator criteria at the time they were first selected (Rose & Wiebe, 2015a). This is in stark contrast to multilateral development banks that have routinely made exemptions to their

selection criteria (Kaja & Werker, 2010; Kilby, 2006). Thus, the most basic and influential effect of the selection indicators is in the “rule based” qualification stage of country selection.

A country’s performance relative to its peers on the WGI was critical to MCA qualification. The WGI wielded a disproportionately large influence on country qualification. This single source made up five of the sixteen selection indicators. In other words, very close to one out of three indicators in MCC’s selection system was derived from Kaufmann and Kraay’s project of aggregating governance data. In the Ruling Justly category, WGI contributed four out of the six indicators, or 67 percent (the other two come from Freedom House’s *Freedom in the World* report). Because countries had to pass at least three indicators in each category, failure to pass the WGI’s four indicators in the Ruling Justly category eliminated a country. Furthermore, a country had to score higher than the WGI Control of Corruption’s median score to qualify irrespective of performance on all other fifteen indicators.

Like the other MCA selection indicators, the WGI indicators are positively correlated with a country’s income levels, as measured by GDP per capita. In theory, this relationship would suggest that countries with higher per capita incomes would have higher than median WGI scores and a greater chance of qualifying relative to poorer countries. As described earlier, some interagency group members were troubled by this income bias. Results of MCA’s qualification process shows that their concern was warranted. Among the low-income group (72 countries per year, on average), qualifying countries have higher GNI per capita on average than countries failing to qualify for every year of selection between 2004 and 2012 (see **Figure 4.11**).<sup>23</sup> The pattern is the same for countries passing and failing the Ruling Justly WGI

---

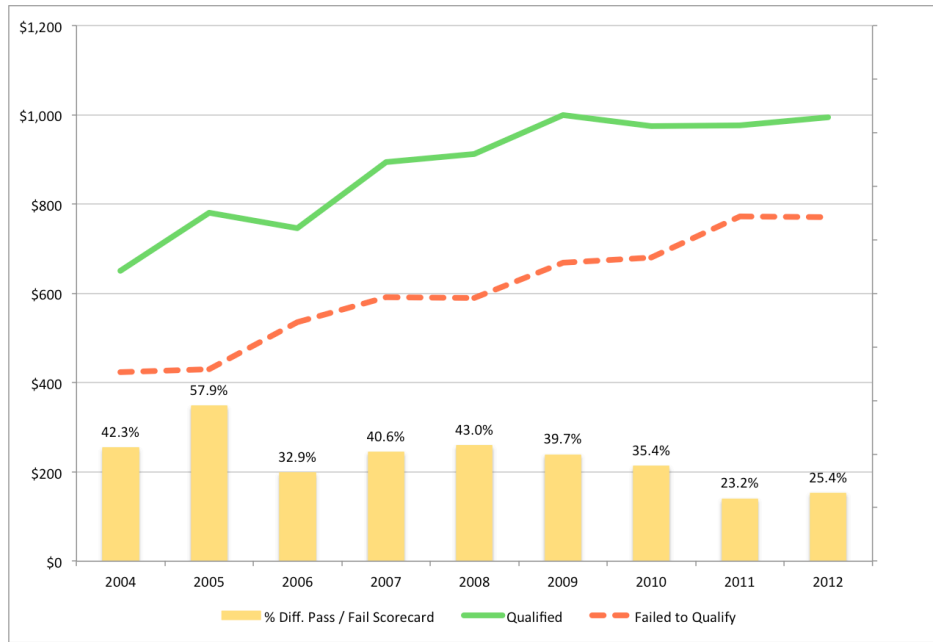
<sup>23</sup> This period was chosen for analysis because in 2013 MCC substantially revised its qualification criteria. The Ruling Justly category changed for the first time. The WGI Voice & Accountability index was substituted with the Freedom of Information index. The new index is a product of Freedom House and the Open Net Initiative. Also, it

indicators (see Appendix B for methodology and data). Countries that score below the median on all four of the WGI indicators—and thus fail to qualify for MCA assistance—have lower GNI per capita, on average as a group, than countries that score above the median on at least one WGI indicator. The differences are substantial. In 2004, the average GNI per capita of countries scoring above the median on at least one Ruling Justly WGI indicator was 47 percent higher than countries failing all four Ruling Justly WGI indicators. On average, across the nine years examined, countries scoring above the median on at least one Ruling Justly WGI indicator had a GNI per capita 34 percent higher than those countries failing all four indicators. In none of the selection years was the average GNI per capita of the group that scored below the median on all four Ruling Justly WGI indicators higher than countries that passed at least one WGI Ruling Justly indicator (see **Figure 4.12**). Because many statutorily prohibited countries, like Myanmar and Zimbabwe, also perform poorly on the WGI and have low incomes, their exclusion from qualification could influence the results. The analysis was repeated without prohibited countries and no significant change in the results was found.

---

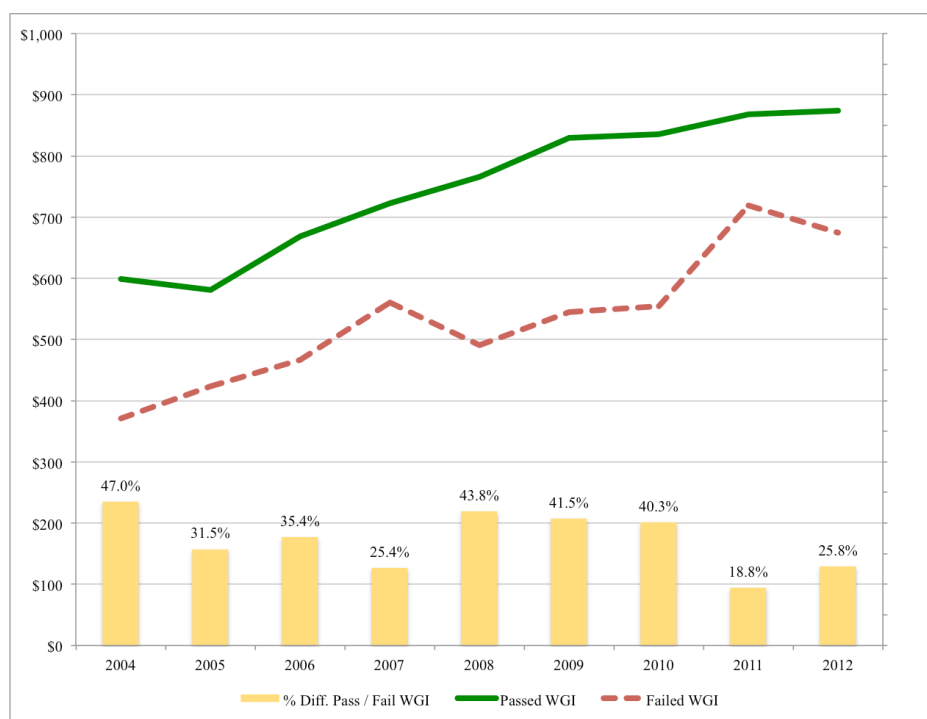
was no longer essential that a country score above the median on at least half of the ruling justly indicators. A country could score above the median on at least half of all indicators, irrespective of which category the indicators were in, and still qualify. They would still need to pass the Control of Corruption hard hurdle and a new hurdle was introduced – a country must score above the median on at least one of the two original Freedom House indicators.

**Figure 4.11.** Differences in GNI Per Capita among Low-Income Countries Qualifying or Failing to Qualify for the MCA



*Source:* Based on author's calculations of MCA performance data. GNI per capita data taken from the World Bank's Worldwide Development Indicators.

**Figure 4.12.** Differences in GNI Per Capita among Low-Income Countries Passing or Failing the WGI



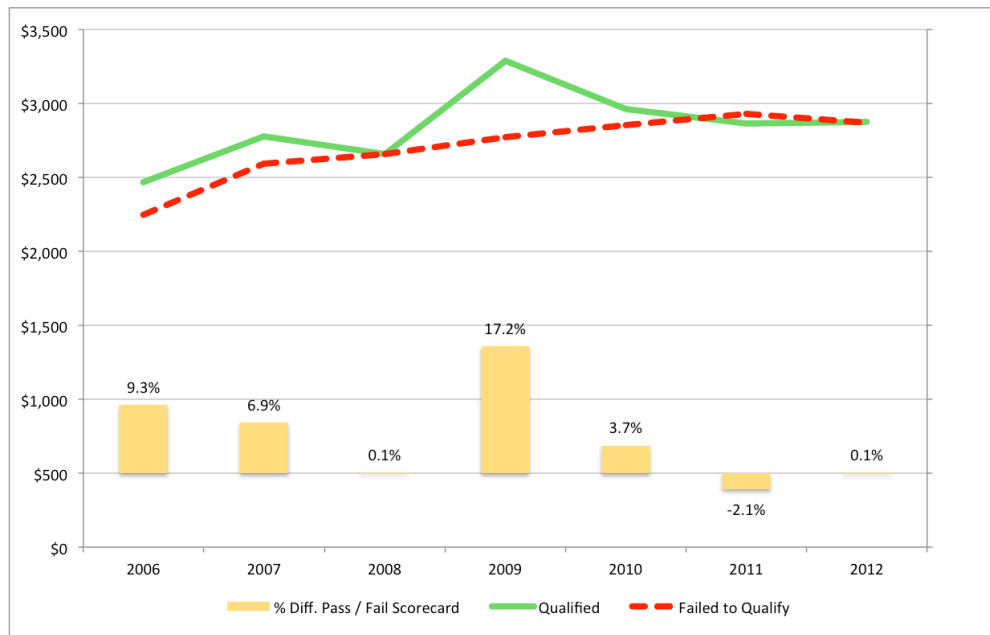
*Source:* Based on author’s calculations of MCA performance data. GNI per capita data taken from the World Bank’s Worldwide Development Indicators.

The story is different for the lower-middle income group. Differences in GNI per capita among this group were far more muted (see **Figure 4.13**). The greatest gap was in the first year these relatively richer countries were considered for MCA assistance.<sup>24</sup> In 2006 the average GNI per capita among lower middle-income countries scoring above the median on at least one Ruling Justly WGI indicator was 11 percent higher than those countries scoring below the median on all four. In general, qualifying lower middle-income countries were not substantially richer than non-qualifiers (see Appendix B, Table B-2). Over the seven-year period analyzed,

<sup>24</sup> The analysis for lower middle-income countries spans seven years, from 2006 to 2012. This is because lower middle-income countries were not considered for assistance until two years after MCA began evaluating countries for funding in 2004.

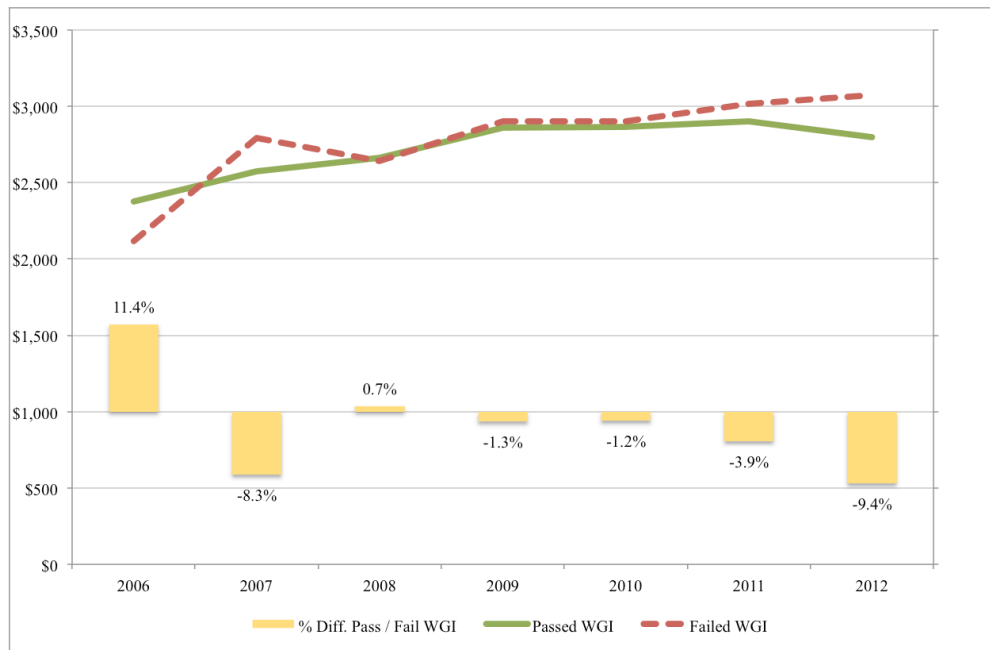
qualifying lower middle-income countries had, on average, a GNI per capita five percent higher than non-qualifiers. Meanwhile, countries that scored below the median on all four Ruling Justly WGI indicators had essentially the same GNI per capita as those that did not (see **Figure 4.14**). Like the analysis for low-income countries, the analysis for lower middle-income countries was repeated with prohibited countries excluded and the results were likewise not significantly different.

**Figure 4.13.** Differences in GNI Per Capita among Lower Middle-Income Countries Qualifying or Failing to Qualify for the MCA



*Source:* Based on author's calculations of MCA performance data. GNI per capita data taken from the World Bank's Worldwide Development Indicators.

**Figure 4.14.** Differences in GNI Per Capita among Lower Middle-Income Countries  
 Passing or Failing the WGI



*Source:* Based on author’s calculations of MCA performance data. GNI per capita data taken from the World Bank’s Worldwide Development Indicators.

The significant differences in income between countries performing better and worse on the WGI in only the low-income country group demonstrate that the positive relationship between governance and income weakens as per capita incomes rise. The Interagency Group’s decision to run separate competitions based on GNI per capita cutoffs mitigated, but did not eliminate, income bias. Income bias was a substantial feature of MCA selection between 2004 and 2012, especially among low-income countries, and the use of governance indicators, and the WGI in particular, contributed to it.

The phenomenon of income bias was not unique to MCC; it was a broader feature of selectivity in official development assistance during this period. While the Bush Administration was striving to establish and operationalize the MCA it was also advocating for a reform strategy in multilateral development banks’ concessional lending operations. Each bank has a fund for its



concessional lending portfolio, such as the World Bank's International Development Association and Asian Development Bank's Asian Development Fund (ADF), separate from non-concessional funding.<sup>25</sup> Capital for these funds comes from wealthier multilateral development bank members such as the United States, France, and Japan (every four years the funds need to be replenished).

The U.S. federal government is a major contributor to concessional programs. This provides the U.S. leverage over how the funds are distributed.<sup>26</sup> As Deputy Treasury Secretary Kenneth Dam said during a speech, "Since we are the biggest contributor to most of these institutions, and since the executive boards use weighted voting, we have considerable voice in whether, when, how much, and under what conditions they lend" (Dam, 2002). The Bush Administration's Treasury Department sought to use this leverage to enact three "sweeping reforms" in multilateral development bank concessional lending: 1) increase grant funding vis-à-vis loans to the poorest countries; 2) link concessional window funding to the achievement of results; and 3) focus more IDA resources on key productivity-driving activities (Replenishment authorizations, 2002). Together with the MCA, this focus on productivity-driven growth at the multilateral development banks was part of the Bush Administration's economic growth agenda: "The President also noted importantly that this approach that we're using in our bilateral assistance is also the same approach that we're using in our so-called growth agenda or reform

---

<sup>25</sup> Concessional lending occurs on much more lenient terms including lower interest rates and longer repayment periods and is reserved for lower income countries or countries that face barriers to private sources of capital. A portion of this window's funds is also distributed as grants that require no repayment.

<sup>26</sup> The US is not as significant a contributor when replenishments are examined relative to contributor countries' GDP. It is also the country that owed the most to the World Bank IDA and "routinely fails to make good on its MDB pledges". In 2002 the US was approximately \$500 million in arrears to global MDBs (Morris, 2017).

agenda for the multilateral development banks, where we're also trying to have a closer marriage between policies and contributions" (White House, 2002: 3).

Emphasis on policies, which included institutions, meant increasing the weight given to governance in the multilateral development banks' internal country assessment formulas (e.g., the World Bank's Country Policy and Institutional Assessment). The Asian Development Bank and African Development Bank were looking to incorporate performance-based allocation systems similar to the Country Policy and Institutional Assessment into their concessional lending programs. Such performance allocation systems were similar in spirit if not form to the MCA country selection system and were meant, like the MCA, to reward countries that had relatively better governance: "These systems provide more resources to those countries that improve governance and take steps to combat corruption, while those who do not take such steps receive fewer resources. For example, under the most recent replenishment of funds in IDA, seventeen countries will have their resource allocations reduced" (Combating Multilateral Development Bank Corruption, 2004).

During negotiations for ADF's eighth replenishment in 2004, the Asian Development Bank's management agreed, at the behest of Treasury and other donor government officials, to place governance at the center of ADF's allocation. The effective weight of governance before the revision was 30 percent. After ADB's board and management revised the system, governance's effective weight increased to 53 percent (ADB, 2004). The ADF's eighth replenishment was a US \$7 billion dollar program to be distributed over a four-year period, from 2005 to 2008, averaging US \$1.75 billion a year (this was equal to Congressional appropriations for the MCA in 2006 and 2007). Asian Development Bank managers anticipated an inequity in funding allocations of this amount as a result of the emphasis on governance performance and,

indeed, the strong performing group ended up receiving, on average, 5.2 times that of the weakest performing group (ADB, 2010). ADB management did not necessarily anticipate, however, that this would also result in an inequality in allocations between relatively poorer and richer countries. They later concluded:

The amount of support to the poorest countries eligible only for ADF funding has declined since the system was introduced... While the amount distributed through the [performance based assessment] system grew by about 22%, differences across country groups are notable. In the first biennial period of ADF IX, \$30 million less per year went to the poorest ADF-only countries. This is equivalent to the average lending program for a country like Kyrgyz Republic or Lao People’s Democratic Republic... In contrast, the large blend countries have benefited greatly under the revised system—their combined annual program gained more than \$200 million per year (ADB, 2008: 4).

In the absence of the revised allocation system, the poorest countries would have received more (see Table 4.3). This meant that funding was shifting away from countries like Cambodia, Kyrgyz Republic, Tajikistan, Nepal, and Laos to countries like Bangladesh, Pakistan, and Vietnam (ADB, 2008).

Table 4.3

*ADF Performance-Based Shares by Country Group, 2002–2008*

Country Group	Allocation Share			Difference from Previous Period	
	2002–2004	2005–2006	2007–2008	2005–2006	2007–2008
Pacific <sup>a</sup>	4.5	4.5	4.5	0.0	0.0
ADF Only	24.9	22.0	21.1	(2.9)	(0.9)
Poorest	21.0	16.9	16.9	(4.1)	(0.1)
Other	3.9	5.1	4.3	1.2	(0.8)
Blend	70.6	73.5	74.4	2.9	0.9
Largest	57.1	60.6	63.7	3.6	3.1
Other	13.6	12.9	10.7	(0.7)	(2.2)
PBA Total	100.0	100.0	100.0	0.0	0.0

( ) = negative value, ADF = Asian Development Fund, PBA = performance-based allocation.

<sup>a</sup> Pacific group includes the set aside for Timor-Leste of \$10 million on average per year since 2007–2008.

Source: Asian Development Bank (2008), *Refining the Performance-Based Allocation of Asian Development Fund Resources*, Manila: Asian Development Bank.

The ADB concluded that “[t]he revisions to the PBA system were designed to strengthen the link between performance and allocation. They created a wide gap between the share of

resources going to the poorest countries and the share going to the large blend borrowers” (ADB, 2008: 5). In negotiations for the ninth ADF replenishment in 2008, the ADB made adjustments to the performance allocation system to counter this decline to poorer countries. The result was a fairer distribution of ADF resources with respect to countries’ per capita incomes (ADB, 2010).

**A Global Underclass.** Unlike the Asian Development Bank’s board and management, the MCC Board of Directors and management never undertook a similar review of potential income bias in its allocations. The only path towards qualification for relatively lower income countries was to leapfrog and sizably outperform their wealthier and already higher scoring peers. This required, at least in part, transforming institutions, economic policies, and human capital investment patterns.<sup>27</sup> Douglass North believed, however, that institutions are path dependent, prone to reproduce themselves, and obstinate to change. A reversal of institutional paths is possible, but usually requires a massive shift in the structure of the state and norms governing the polity: “While the rules may be changed overnight, the informal norms usually change only gradually. Since it is the norms that provide ‘legitimacy’ to a set of rules, revolutionary change is never as revolutionary as its supporters desire and performance will be different than anticipated” (North, 1994: 366).

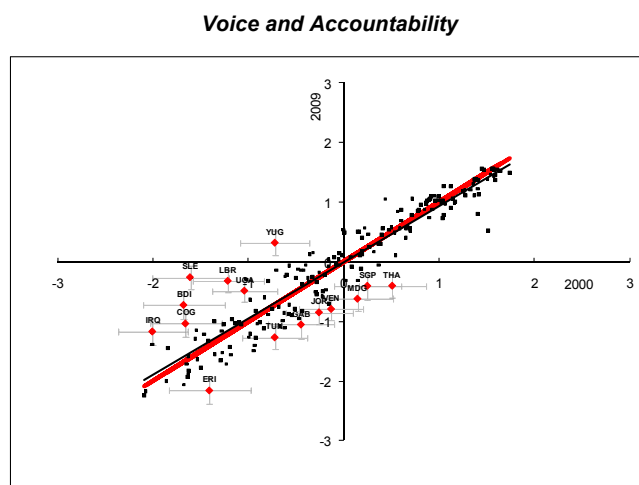
Countries’ WGI scores over time appear to corroborate North’s conclusions. Based on an analysis of countries’ changes in WGI scores over a one-decade period, 2000 to 2009, roughly equivalent to the analysis of MCA qualification outcomes above, Kaufmann and Kraay concluded that “changes in our estimates of governance in most countries are relatively small” (Kaufmann et al., 2011: 234). In fact, over this period, less than ten percent of all countries in

---

<sup>27</sup> Since governance indicators are based on perceptions, one could presumably just run a campaign to alter the impressions of a country’s institutions among the diverse actors populating the polls and surveys that measure governance but this could also be challenging, costly, and lead to accusations of “gaming the indicators.”

the WGI database recorded a statistically significant change in governance scores (Kaufmann et al., 2011). Moreover, not all these changes were improvements. Changes were about equally divided between improvements and deteriorations in scores setting up a zero-sum competition between countries, which is a feature of indicators like the WGI (Espeland & Sauder, 2007). Kaufmann and Kraay graphically represented this phenomenon using scatterplots of countries' scores accompanied by two trend lines (see **Figure 4.15**).

**Figure 4.15.** Scatterplot Comparison of Countries' 2000 and 2009 WGI Voice and Accountability Estimates



*Source:* Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2), 220–246.

*Notes:* For each indicator, the trend lines almost completely overlap, demonstrating that statistically significant changes over this period were rare.

Limited changes in WGI scores over time mean that initial scores risk locking countries into either a qualifying or non-qualifying track—the same countries will keep qualifying or failing year after year. Because WGI performance is correlated with income, this means that over time relatively richer countries will keep qualifying for funds and relatively poorer countries will keep failing to qualify. Low-income countries such as Cameroon, Togo, Chad,

Mauritania, Eritrea, Djibouti, Cambodia, Laos, Papua New Guinea, among others, failed to qualify in every year between 2004, the first year of the MCA competition, and 2012.

Meanwhile MCC has had a hard time finding new qualifying countries. The agency has resorted to signing additional compacts with countries that already received them because so few new countries qualify. This is not something the agency ever envisioned doing (Rose & Wiebe, 2015a).

These and other similarly “poorly” performing countries represent a new global underclass. The MCA interagency group elevated conservative philosophies around poverty in the US to a global scale. While the economic policy community’s efforts reversed an accelerating downward trend of US bilateral ODA during the 1990s, its strategy ensured that the money would only go to those countries governance indicators deemed deserving. Countries that failed to measurably demonstrate a culture of “self help” through improvements in governance were deemed undeserving (Katz, 2013).

## **Conclusion**

This chapter described how following the 27th G8 Summit in Genoa, the Bush administration found itself in a conundrum. At the international level, it was faced with calls to increase development financing for relatively lower-income countries. Domestically, it faced rising skepticism around the benefits of foreign aid, particularly by conservatives who happened to occupy many White House offices, key Cabinet posts, and control Congress. It described how NSC officials came to settle on selectivity—spending foreign aid in countries where it will be most effective—as a strategy for simultaneously increasing development finance budgets and curbing criticism based on World Bank econometric research. They knew little, however, about exactly how to select these countries.

The research presented here showed how a group of Treasury economists, led by Steven Radelet who had strong ties to World Bank economists and the Center for Global Development, solved the NSC's selectivity challenge. In doing so, economic experts proactively presented a solution—targeting countries based on “objective” and “transparent” governance (among other) indicators—that put them in charge of shaping the details of the MCA country selection system. The NSC, in turn, gained more freedom to move forward with its new aid initiative in an unreceptive environment. This chapter also described how the road to selectivity by indicators was not an entirely smooth process. Treasury economists and their economic policy partners in the NSC faced opposition from the foreign policy community and had to engage in the kind of political work often believed to be the reserve of politicians to overcome resistance.

Finally, the chapter presented evidence of the political ramifications of a selectivity strategy based on World Bank governance indicators. The WGI did not eliminate bias in the selection of countries for funding as the economic policy community claimed, but rather introduced the biases of an assemblage of actors such as survey instrument writers and responders, World Bank research economists, and Treasury economists into bilateral aid decisions. These new biases affected how decisions were made, such as the dismissal of countries' unique institutional circumstances, and funds were distributed, such as disfavoring relatively lower-income countries.

## Chapter 5: The Power of Growth Diagnostics in Development Planning

### **Introduction**

This second case examines the Millennium Challenge Corporation's (MCC) use of growth diagnostics to determine what is included in multi-million dollar development grants to country governments. Three Harvard university economists devised a decision-tree framework and a set of data-driven empirical tests to determine what specific constraints limit a country's level of private investment. This approach was a reaction to dissatisfaction with the shortcomings of prescriptive structural adjustment programs. MCC began using growth diagnostics after having already developed nearly 20 bilateral aid agreements, or compacts, with partner countries in a more ad-hoc fashion and relying on existing planning vehicles such as Poverty Reduction Strategy Papers. This chapter describes how MCC economists were able to adapt growth diagnostics for compact development purposes and why the agency eventually adopted the method. It discusses how the rationale driving growth diagnostics was at odds with a more political-economic and social justice oriented approach to development and what form this tension took inside the organization. It goes on to compare the process and outcomes of MCC's earlier compact development process and outcomes with the process and outcomes for developing 14 compacts using growth diagnostics. The chapter closes by providing evidence of how growth diagnostics contributed to changes in organizational power relations and outcomes such as the distribution of funding across sectors, space, and project types.

### **Growth Diagnostics in Perspective**

**Origins of Growth Diagnostics: Harvard's Kennedy School.** Growth diagnostics is a methodology pioneered by three economists—Ricardo Hausmann, Dani Rodrik, and Andres Velasco—at Harvard University's John F. Kennedy School of Government. They continued a



long tradition of Harvard University involvement in development issues. Rodrik is the most well-known due in part to having published a popular book titled *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth* (Rodrik, 2009). He has a doctorate in economics from Princeton University and is the only one of the three to have not served as a senior government official. Rodrik is currently Ford Foundation Professor of International Political Economy at the Kennedy School. Ricardo Hausmann has a doctorate in economics from Cornell University and is the Professor of the Practice of Economic Development and Director of the Kennedy School's Center for International Development. He was previously the Inter-American Development Bank's chief economist and Venezuela's Minister of Planning from 1992 to 1993. Andres Velasco, also a Latin American, has a doctorate in economics from Columbia University and was Professor of Development and International Finance at the Kennedy School from 2000 to 2011. He was Chile's Finance Minister from 2006 to 2010.

Harvard's involvement in contemporary international development can be traced back to the Development Advisory Service, also known as the Harvard Advisory Group, which Edward S. Mason established in 1962 at the then Harvard Center for International Affairs. The Advisory Group consisted of a small circle of approximately ten economists that counseled newly independent countries on economic planning (Gordon, 1963). The Harvard Advisory Group would later be converted into the Harvard Institute for International Development (HIID). Dwight Perkins was HIID's first president. Jeffrey Sachs, an important figure in contemporary debates in development economics and an influential policy advisor to the United Nations, took over in 1995. The HIID produced over 700 Development Discussion Papers and a number of policy analysis products such as the Global Competitiveness Report, a methodology focused on firms' and enterprises' impressions of country governance and operating conditions. The HIID

was dissolved in 1999 and Sachs formed the Center for International Development at the Kennedy School. The Center for International Development is now Harvard's hub for international development and the institutional home for Rodrik, Hausmann, and Velasco (HRV). They all teach in its Masters in Public Administration / International Development (MPA/ID) degree program. In conjunction with the Kennedy School, the Center for International Development created the MPA/ID to train those seeking careers in international development. The faculty "was dominated by eminent development macroeconomists" and the degree focused on "PhD level economic theory" and "hard mathematics" (Blattman, 2008). The Center for International Development is also focused on research. It has solicited over 300 working papers from across Harvard, conducts policy evaluations, and established the Growth Lab where growth diagnostics have been conducted for at least 10 countries.

**Growth Diagnostics and the (Post-) Washington Consensus.** In 1989 John Williamson, a senior fellow at the Peterson Institute for International Economics (a Washington-based think tank) who coined the term Washington Consensus convened a conference with reformers from ten Latin American countries at the Peterson Institute to "convince a skeptical Washington that policies were indeed changing" (Williamson, 2003). The Washington Consensus was originally a synopsis of policy reforms introduced by economists and economic policy officials from these Latin American governments. The Consensus was adopted and adapted by influential development institutions such as the International Monetary Fund (IMF), World Bank, United States Treasury (Treasury), and Inter-American Development Bank (IADB). Rightly or wrongly, Williamson's synopsis became synonymous with these institutions' approach to structural adjustment lending and policy programming that included administrative requirements of conditionality that made funding disbursements contingent on

country governments executing a set of economic, policy, and institutional reforms.

The Washington Consensus approach to policy reform for growth and development resulted in widespread disappointment and in some cases outright protest. These reforms aimed at macroeconomic stability and market liberalization were explicitly directed at reviving growth but implicitly geared towards resolving the Third World debt crisis that emerged at the end of the 1970s (Vasquez, 1996). Evidence would later show that reforms were more successful at ensuring debts were repaid than reviving growth. The 1980s would come to be known as Latin America's "lost decade", the Washington Consensus approach to development in Sub-Saharan Africa actually retarded growth, and the IMF recommendations to eliminate capital controls in South East Asia exacerbated the Asian financial crisis (Van De Walle, 1999).

Hausmann, Rodrik, and Velasco developed growth diagnostics as a direct response to the embattled Washington Consensus set of policy reforms popular in the 1980s and 1990s. Many economists, including proponents of a growth diagnostic approach to reform, agree that the ten reforms laid out by Williamson can go a long way towards spurring growth.<sup>28</sup> Their problem is less with the specific reform program *per se* than the way a set of reform experiences from select Latin American countries became a standardized prescription for policy reform across the globe (Hausmann, Rodrik, & Velasco, 2004; Headey, 2009). Hausmann, Rodrik, and Velasco also felt that the program tried to do too much all at once without any process of comparing or prioritizing reforms. Dani Rodrik described it as "doing as much reform as you can, the best as

---

<sup>28</sup> This reform program argued that: 1) Budget deficits should be small enough to avoid inflation; 2) public expenditure should be allocated efficiently; 3) taxes should be drawn from a broad base and cut when possible; 4) interest rates should be determined by liberalized financial markets; 5) exchange rates should be unified to increase exports; 6) trade restrictions should be lifted and tariffs lowered; 7) barriers to FDI should be abolished; 8) state enterprises should be privatized; 9) regulations that restrict market entry or competition should be eliminated; and 10) private property right should be established and secured (Williamson, 2004).

you can” (Rodrik, 2009). The diagnostic approach aims to select the “binding constraint” to growth and is meant to take a country’s context, or unique economic environment, into account. It also prioritizes among multiple constraints by estimating the “growth dividend” derived from the removal of various constraints. Rodrik stated, “Rarely will the advisor ask whether the problem at hand constitutes a truly binding constraint on economic growth, and whether the long list of institutional reforms on offer are well targeted at the economy’s present needs. But governments are constrained by limits on their resources ... They have to make choices” (Rodrik, 2009: 5).

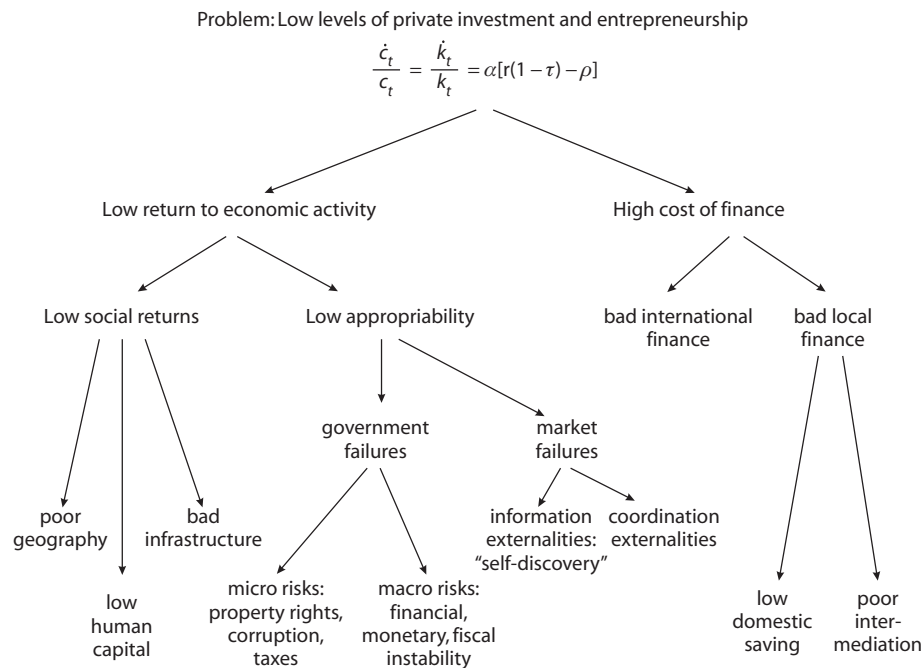
### **The Mechanics of a Growth Diagnostic**

**The Decision Tree.** The growth diagnostic relies on a decision tree framework. A decision tree is a pre-structured decision-making template that allows organizational actors to structure choices, specify alternatives, and identify outcomes all while respecting the logic of causality (Cabantous et al., 2010). As an analytical tool, the decision tree has been around for decades and is not unique to growth diagnostics. Its roots can be found in the field of decision analysis that emerged in the 1960s as a discipline distinct from decision theory, system modeling, and operations research. Decision analysis is built on two main foundations. The first is the Bayesian School of probability. The second is the economic approach of utility measurement (Headey, 2009). According to Rodrik, the growth diagnostic “sketches a systematic process for identifying binding constraints and prioritizing policy reforms in multilateral agencies and bilateral donors” (Rodrik, 2009: 35). How a decision tree is populated is important. A decision tree starts off “naked”. It gets “dressed” as theory is added to guide questions and data is added to answer them (Cabantous et al., 2010). While Rodrik emphasizes that diagnostics require eclecticism in both the use of theory and evidence and have no room for

dogmatism, ultimately the theories and data chosen for the exercise express a distinctive rationale for what is possible by how they define problem(s), diagnose causes, and identify solutions.

The ultimate goal of growth diagnostics is to increase the rate and level of growth *via* private investment and entrepreneurship. The growth diagnostic decision tree heuristic is used to diagnose the problem of low private investment by conducting a battery of data-driven, empirical tests. **Figure 5.1** illustrates the growth diagnostic framework.

**Figure 5.1.** Growth Diagnostic Decision Tree



Source: Hausmann, R., Rodrik, D., & Velasco, A. (2008). Growth diagnostics. *The Washington consensus reconsidered: Towards a new global governance*. Oxford: Oxford University Press.

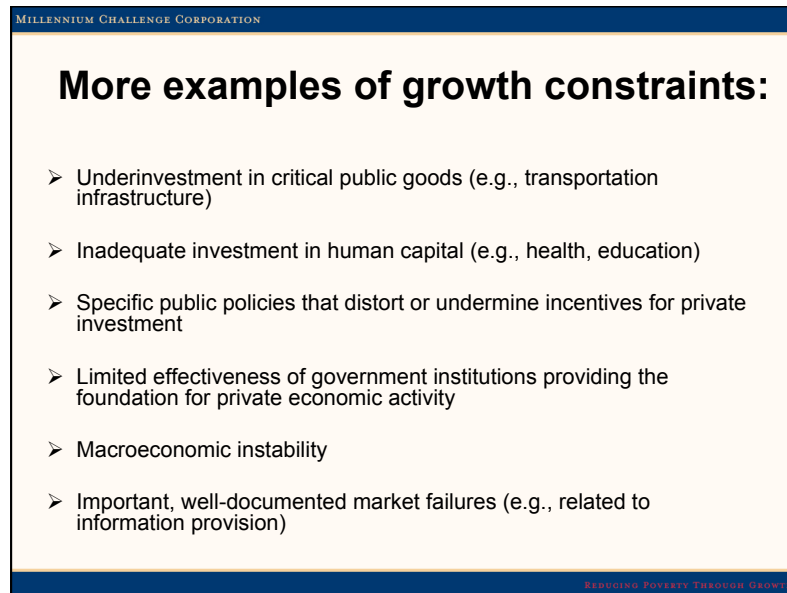
In the decision-tree represented in **Figure 5.1**, the “High cost of finance” branch represents constraints that are related to relatively high interest rates that prevent returns on investment and low capital accumulation rates that in turn stunt growth. This can be due to

either “bad international finance” where the country is too “risky” and is having trouble attracting capital from outside or “bad local finance” where the risks of moral hazard and adverse selection increase or savers are not being linked to borrowers (i.e., poor intermediation). The other major branch of the tree—“Low return to economic activity”—represents constraints that are less about attracting capital than generating “expected” private returns on that capital, or enough profits. Under that branch, “Low social returns” refers to: i) insufficient human capital (e.g., skilled people) to generate returns on an investment; ii) inadequate infrastructure, such as electricity or roads, that makes investing more costly and therefore drives down profits; and iii) poor geography (e.g., isolation) as exhibited by high logistical transaction costs that hamper profits.

The other constraint to growth is “Low appropriability” or the capacity of the firm to maximize profits from its added value. One aspect that limits appropriability is government failures. These are viewed as actions actively taken or avoided by the government that affect firms’ profits (e.g., bribes, poorly defined property rights, or high taxes). The others are government failures in the macro economy such as high national debt and poor monetary policy, both of which can lead to inflation or currency depreciation and affect firms’ savings. Low appropriability can also be affected by market failures such as the inability to identify and generate productive activities within a particular context.

A slide from an MCC presentation (see **Figure 5.2**) from its Economic Analysis department shows what the agency believes to be some common examples of growth constraints

**Figure 5.2.** MCC Examples of Growth Constraints



*Source:* Wiebe, F. (2007, March 19). *Economic Analysis in the MCC*. Presentation delivered at MCC-University. Washington, DC: Millennium Challenge Corporation.

Multiple schools of economic thought are captured in **Figure 5.2**. This is something that growth diagnostics’ creators advocated and hoped for (Headey, 2009). The first two bullet points on investment in infrastructure and human capital represent a renewed role for public provision of public goods. The third and fourth bullet points represent recent thinking on institutions and their importance to growth and development. The bullet point on macroeconomic stability represents the continued emphasis on approaches important to promoters of the Washington Consensus.

To narrow-down possible constraints to growth, analysts eliminate constraints they feel do not apply a particular country’s case based on prior research or analysis and then conduct tests on the remaining candidates. Ricardo Hausmann and others at Harvard published a guide to doing growth diagnostics called a “Mindbook” where they outlined four diagnostic tests. As they put it, “No single test or symptom is clearly definitive...A differential diagnosis requires the application of several tests, which are collected and aggregated in the proper Bayesian

framework” (Hausmann, Klinger, & Wagner, 2008: 31). The four tests for each potential constraint analyze: 1) the constraint’s (shadow) price; 2) the relationship of the constraint to investment and growth levels; 3) private actors’ response to the constraint; and 4) how “intensive” actors are in the constraint (Hausmann et al., 2008). Each of these will be discussed in detail later.

### **MCC Compact Development: The Early Years**

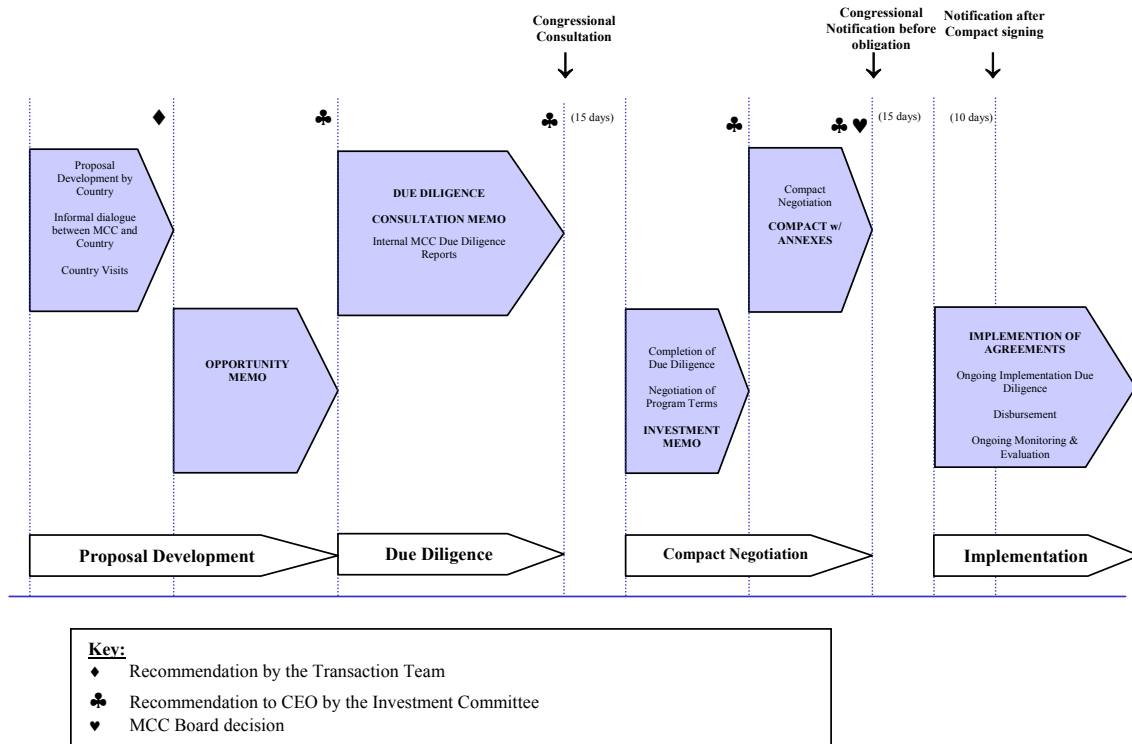
MCC began adopting growth diagnostics for compact development, or designing bilateral investment treaties with partner country governments, about half a decade after its inception in 2009. Before MCC began using the growth diagnostic, the agency designed nearly twenty bilateral investment agreements, or compacts. Early compacts were developed in an ad hoc way among MCC senior officials, sector specialists, and a partner government’s “core team” comprised of officials from the Prime Minister’s office and various ministries. Country proposals for compact funding were often informed by or directly based on Poverty Reduction Strategy Papers (PRSP), a vehicle for donor assistance governed jointly by the World Bank, International Monetary Fund, and the country government using varying degrees of civil society input. Senior officials and MCC economists were uneasy with this informal approach to compact development, albeit for different reasons, that produced compacts with many disjointed parts from country proposals full of “pork barrel” spending. Economists also felt that MCC-sector officials were not objective in their negotiations with the core team and steered projects towards their own sectors.

**Planning Without Diagnostics: “Making the Sausage.”** Prior to growth diagnostics, or constraints analyses (CA) as MCC called them, there were four general parts to the compact development process: country visits, project development, compact negotiation, and project



appraisal, or “due diligence” (see **Figure 5.3**). Different actors in the organization played different roles in these processes. Country team leaders, senior officials, and sector specialists dominated country visits. An MCC delegation consisting of the CEO and/or other senior officials such as Vice Presidents and senior sector specialists visited their core team counterparts—a group of senior ministry and other government officials set up by the country’s President or Prime Minister. It was common for trips to take place every two to three months over a 1 to 2 year process. In general, MCC senior officials dominated early trips.

**Figure 5.3.** Early Compact Development Process



Source: MCC (2005b). *Working with MCC: Overview of Steps and Processes*. Washington, DC: Millennium Challenge Corporation.

The initial trip was usually a diplomatic formality where the delegation would describe how MCC working principles were similar to or different from other United States (US) government agencies and international donors. Sector specialists joined on subsequent trips. On these trips and the periods in-between, the outline of the proposal would take shape through iterative, interactional project development negotiation. In later trips, other technical staff from Economic Analysis, Monitoring and Evaluation (M&E), and Environmental and Social Assessment (ESA), usually after the outlines of a proposal had been developed, would visit their counterparts. The technical staff's trips were usually focused on identifying and collecting data for various project appraisal tasks such as identifying monitoring indicators, calculating economic rates of return (ERR), and developing environmental and social impact assessments. The economists would produce "Growth Reports". Growth reports focused on the macroeconomic situation in a country and how it might affect the compact. A growth report for Cape Verde asked questions such as "Are the 'consensus' views about growth constraints accurate and/or backed up by evidence?" (Warner, 2004: 2). In answering that question, the report stated that "[m]any infrastructure deficiencies are also identified as 'constraints'; the problem is that there is much less clarity about which of these are economical to remove...The local [core] team is well capable of identifying problems but seems to be trying to satisfy a donor checklist rather than providing any sort of growth analysis from which project priorities could be rationally derived" (Warner, 2004: 2). These growth reports foreshadowed economists' dissatisfaction with country proposals.

Following the initial visits the core team was officially invited to submit a Compact Proposal outlining its development program. Neither the exact compact amount nor content was known at that time, although the core team often had a general idea of what they were going to

include. As compact development progressed the core team liaised with the MCC Transaction Team—a group of more junior technical and sector officials from sectors, ESA, M&E and other technical divisions led by a Transaction Team Leader. There was regular telephone and electronic communication that included the sharing of documents, spreadsheets, and other program or project level data. This process was often conducted under intense time pressure. One senior official referred to the process as a “SWAT team” approach, referring to the specially equipped and trained United States (US) law enforcement units reserved for high intensity and risky situations (Interviewee 2H, personal communication, March 2, 2016).

Negotiating a compact was a process that involved the core team deciding which projects to propose and the transaction team deciding which projects to fund. It was sometimes a downright disorderly process (Interviewee 2B, personal communication, February 11, 2016). Beyond MCC’s baseline requirements for environmental and social protection for things like endangered species and displaced populations, technical feasibility in the form of engineering guidelines, and economic rates of return, there was no structured decision-making framework.<sup>29</sup> There was considerable flexibility in what a country could propose and advocate for and limited guidance on what MCC would ultimately accept. In the words of one senior MCC official, “In trying to come to some sort of agreement in the country, we had no good analytic tools for asking ‘does it make sense?’” (Interviewee 2P, personal communication, March 29, 2016).

Countries would propose projects to be financed by MCC using the rationale they desired and proposals were often based on national development plans called Poverty Reduction Strategy

---

<sup>29</sup> A project’s economic rate of return (ERR) had to surpass a “hurdle rate” equal to two times the average growth rate in a country’s GDP over the most recent three years of data—but this was usually an afterthought. This formulation was amended in 2007 to include a minimum rate of 10 percent, applicable to slow-growing countries, and a maximum of 15 percent, to “protect” faster growing countries. The current hurdle rate is 10 percent irrespective of the country (Rose & Wiebe, 2015a).

Papers (PRSP). Local participation was an important component of compact development. The backlash against structural adjustment's heavy-handed, top-down development approaches spurred new governing philosophies among development finance organizations (Best, 2014). This reformed approach was pioneered through the World Bank's Poverty Reduction Strategy (PRS), the mechanism that facilitated the creation of PRSPs. At the time MCC was created, 28 countries had already produced a PRSP and another 45 countries were in the process or near completion (Bretton Woods Project, 2003). The PRSP outlines a national program for poverty reduction that is the foundation for lending programs with the World Bank and IMF.<sup>30</sup> Six core principles underlie the development of poverty reduction strategies. Strategies should be: 1) country-driven; 2) results-oriented; 3) participatory; 4) comprehensive; 5) partnership oriented; and 6) focused on the long-term (Piron & Evans, 2004). By "country-driven" and "participatory" the designers believed that strategies should represent a consensual view of what actions should be taken based on broad-based participation by all relevant stakeholders, including civil society institutions.

The content of PRSPs was meant to serve as the assistance framework for external partners, including bilateral donors like the MCC. This was referred to as donor alignment in the development assistance community. Alongside multilateral development banks (MDB), the United States government and its major bilateral development agencies adopted the poverty reduction strategy's principles of donor alignment and called it "Country Ownership". Country ownership means that partner country governments should exercise leadership in the design and

---

<sup>30</sup> Countries looking for debt relief under the Heavily Indebted Poor Countries initiative were required to produce a PRSP that would then require approval from the World Bank and IMF boards. The World Bank and IMF developed the PRSP approach as a response to evident weaknesses in relations between poor countries and the Bretton Woods Institutions—in particular a lack of focus on poverty and no country ownership of reforms (Bretton Woods Project, 2003).

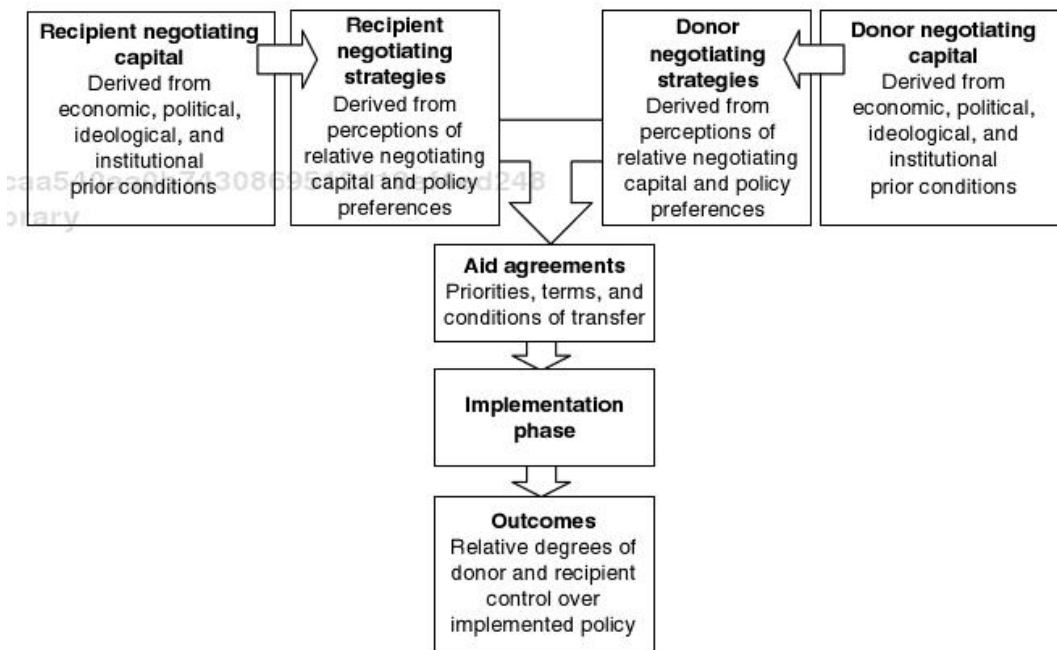
implementation of their own national development strategies with meaningful participation from local stakeholders. In particular, bilateral agencies should support the partner country's leadership in these roles by closely aligning with local priorities (Rose, Kalow, Parks, & Masaki, 2016). The MCC always consulted a country's PRSP, if available, and other national or sectoral planning documents when reviewing country proposals and developing compacts. Among MCC's first 19 compacts, 13 had some basis in the PRSP. The other six had their basis, at least in part, in national development plans that usually included a poverty reduction focus and priorities similar to those found in PRSPs.

Because PRSPs and these national plans were broad and, as some have critiqued, lacked focus or an emphasis on priorities, there were multiple ways a compact could be formulated and designed and still fit within the PRSP framework. Some countries would use its PRSP as a rhetorical device while others would align their proposals very closely with ongoing PRSP or national planning exercises (Interviewee 2T, personal communication, April 5, 2016). Furthermore, MCC's legislation required compact-eligible countries to conduct consultative processes specifically for compact development. Like the PRSP consultations, these varied greatly from country to country and MCC's experiences with them will be discussed later. Thus, MCC's project definition phase of compact development was an iterative process of development diplomacy and negotiation.

Once an initial proposal was submitted, negotiations would begin and outcomes were contingent on context and strategy. According to Whitfield (2009): "The outcomes of aid negotiations are the product of the encounter between recipient and donor preferences...the ability of each actor to successfully achieve their preferred outcomes is heavily constrained by the conditions under which each faces the other and the negotiation strategies they adopt"

(Whitfield, 2009: 38). In the negotiation process at MCC, some of the key variables dictating whether MCC or the country would hold more sway included the specific transaction and core team members, diplomatic relations between the US and the partner country, and contextual factors specific to that country’s planning environment. These factors determined a country’s “negotiating capital” (see **Figure 5.4**). One senior sector specialist remarked, “Sometimes MCC would exercise a heavy hand...sometimes governments would have much more of a say in what was going to be funded” (Interviewee 2B, personal communication, February 11, 2016).

**Figure 5.4.** Simplified Model of an Aid Negotiation



Source: Whitfield, L. (2009). *The Politics of Aid: African Strategies for Dealing with Donors*. Oxford University Press.

Another senior sector specialist compared these early compact negotiations at MCC to how legislation is made on Capitol Hill: “I know there were interests from many sides, and those interests coalesced around certain options. I won’t tell you that there was no influence from the President or other political angles, but ... politics is making the sausage. And the sausage

sometimes was made in ways that we didn't like too much, but it had to be made. Otherwise things would not move forward" (Interviewee 2Q, personal communication, March 29, 2016).<sup>31</sup>

The unstructured and informal nature of the compact development process made the economists and some senior officials uneasy. There were three main concerns around country proposals in the early years. The first was that the most powerful ministries in-country were dominating the process and presenting projects they favored. One senior official rhetorically asked, "Politically important, politically connected, powerful people running the ministries, they get the largest donor financed projects, and guess whose projects show up first when MCC asks countries to propose stuff?" (Interviewee 2J, personal communication, March 15, 2016).

The second main concern was that countries were proposing projects that other donors rejected, often because they didn't make sense or were of dubious quality. MCC officials, curiously, referred to these projects as "dogs": "It wasn't prevalent, but there were some countries that just threw dogs at MCC... What happened was some countries were very cynical about [the process] and just started throwing projects out there, some that had been rejected by the World Bank or USAID or by other big donors" (Interviewee 2J, personal communication, March 15, 2016). According to a senior official, "[the core team] would put out... a project that had been sitting on the shelf for a while, in some cases it hadn't been done because it made no damn sense or because it was too complicated" (Interviewee 2P, personal communication, March 29, 2016). This phenomenon was not widespread but it occurred often enough to frustrate some senior MCC officials. It often fell to the transaction team leader, sector leaders, and their respective core-team counterparts to sort it all out and narrow the focus of the proposal.

---

<sup>31</sup> For an in-depth discussion of how some aspects of this negotiation unfolded see *The Millennium Challenge Corporation and Ghana* (Ebrahim & Rangan, 2009).

This led to the third main concern. Economists grew worried that sector specialists were engaging with the core team and “drumming up business” for their respective sectors (Interviewee 2F, personal communication, February 23, 2016). According to one economist, on trips to countries sector specialists would assure their core team counterparts that their sectors or projects would be represented in the compact. According to another MCC economist, “Very often [compact proposals] have contained a multitude of activities across several sectors, with no clear connection to the country’s key impediments to growth...MCC often felt compelled to invest resources in such proposals due to the time and political capital invested by the sponsor or preferences of MCC sector specialists” (Anderson, Breitbarth, & Osborne, 2010: 2). They were also concerned about MCC funding projects resembling the proverbial “bridge to nowhere” (Rose & Wiebe, 2015b). Economists accused sectors of engaging in “selection bias” that would lead to “pork barrel” spending (Warner, 2004). They believed that pork barrel spending, in addition to being inherently wrong and inefficient, also detracted from economic growth and ultimately hurt the poor: “the political priority to use MCC grant funds may be very high but may be entirely unrelated to the most pressing problems that limit investment and growth in the country as a whole” (Rose & Wiebe, 2015b: 8).

**Summary of Early Compacts.** This interactionist and iterative negotiation process led by senior officials and sector specialists resulted in MCC’s first set of 19 compacts signed between 2005 and 2009. These compacts provided, on average, US \$401 million to each selected country (all dollar figures throughout are in US 2015 constant terms). Among these early compacts, the median compact provided US \$354 million because a few compacts like Tanzania’s (\$773 million) were much larger than others. MCC committed over \$7.2 billion in total across the 19 compacts. These compacts represented 212 unique activities. The typical



compact had three projects with each project having between four and five activities (see Table 5.1). Projects focused on an objective in a particular sector. For example, a transport project might be oriented around reducing travel times and costs in a particular region of the country. The activities might include the construction of new roads, rehabilitation of existing roads, and the establishment of a road maintenance and safety fund.

Table 5.1

*Basic Profile of First Nineteen Compacts, 2005–2009*

	Mean	Median	Min	Max	Total
Number of Sectors	4.3	4	1	7	18
Number of Projects	3.3	3	1	5	57
Number of Activities	13	14	2	19	212
Compact Funding (\$)	401	354	76.1	787	7,218
Project funding (\$)	123	104	8.6	413	6,978
Activity funding (\$)	32.9	12.7	0.12	390	6,978

*Source:* Compacts' Summaries of Multi-Year Financial Plans.

*Notes:* All dollar figures are in 2015 millions of dollars committed. Project and compact funding totals differ because compact funding includes general administrative and management costs. First 19 compacts include those signed between 2005 and 2009: Armenia, Benin, Burkina Faso, Cape Verde, El Salvador, Georgia, Ghana, Honduras, Lesotho, Madagascar, Mali, Mongolia, Morocco, Mozambique, Namibia, Nicaragua, Senegal, Tanzania, Vanuatu.

Eighteen sectors were represented in the first set of compacts. Among these 18 sectors were a total of 58 sectoral sub-activities. The sector with the highest level of investment was Transport, followed by Agriculture, Water and Sanitation, and Land and Property Rights. These four sectors make up just over three-fourths of all funding among these compacts. The remaining 13 sectors make up 24 percent. Exact figures and percentages are displayed in Table 5.2.

Table 5.2

*Sector Distribution of First Nineteen Compacts*

Rank based on funding	Sector	Funding (\$)	Percent (%) of Total
1	Transport	2,924,398,304	41.9
2	Agriculture	1,555,610,617	22.3
3	Water & Sanitation	496,281,544	7.1
4	Land & Property Rights	323,866,864	4.6
5	Energy	287,592,460	4.1
6	Education	251,747,195	3.6
7	Other - Multisector	221,744,696	3.2
8	Banking & Financial Services	172,034,161	2.5
9	Health	157,338,077	2.3
10	Fishing	131,092,974	1.9
11	Industrial Development	126,306,371	1.8
12	Social Infrastructure & Services	110,999,082	1.6
13	Business Development Services	83,858,836	1.2
14	Tourism	75,219,899	1.1
15	Education, level unspecified	37,014,052	0.5
16	Forestry	15,872,319	0.2
17	Government & Civil Society	5,923,527	0.1
18	Reproductive Health	1,513,547	0.0
	Total	6,978,414,526	100.0

*Source:* Compacts' Summaries of Multi-Year Financial Plans.

*Notes:* All figures in 2015 dollars. Total does not include funding for general management and administration. Sector and sub-activity definitions are taken from the Organization for Economic Cooperation and Development's (OECD) Development Assistance Committee (DAC).

While compacts with three or four projects or sectors and a dozen or so activities were typical, some compacts, such as the Morocco's first compact, had as many as seven distinct sectors. The Lesotho Compact had nineteen activities across five sectors. Compacts with several sectors and many different, sometimes disconnected, projects and activities came to be known as "Christmas Trees". One transaction team leader said, "We came up with Christmas tree

compacts which are hellish to implement. A Christmas tree because you get seventeen different activities in three different sectors, some of which hang together, some of which don't, but it's what you get from the universal hundred things [the core team] wanted" (Interviewee 2R, personal communication, March 30, 2016). The many activities of a compact represented the different ornaments on the tree.<sup>32</sup> Most involved in compact development grew frustrated that after all this work a proposal still resulted in a Christmas tree compact. MCC senior managers felt that the negotiation process of winnowing down proposals from long laundry lists (as early proposals were often called) was inefficient and wearisome. As a new agency with limited staff, the MCC implementation teams found it taxing to manage such compacts, especially within the mandated five-year implementation timeline. The Christmas metaphor had another meaning as well—many of these compact projects and activities were more like gifts than thoughtful economic investments. Some MCC officials and many economists believed that the different projects and activities were gifts to different ministers, government agencies, or sector interests.

### **Stabilizing Constraints Analyses**

**The Pitch: Constraints Analyses for “Enhanced Engagement.”** Senior managers were looking for a solution to what was a laborious compact development effort—they wanted a way to quickly winnow proposals down into compacts. MCC economists pitched growth diagnostics, or constraints analysis, to bring an “objective” framework to this “politicized” process. They had a solution to MCC’s laundry list proposals, Christmas tree compacts, and laborious and ad-hoc compact development process. A report by MCC’s economic analysis team put it thus: “These experiences highlighted the need for an objective framework to focus compact proposal

---

<sup>32</sup> American Christmas trees are known for their bountiful and often quirky ornaments, usually gathered over many years from various friends and family members. While they all get added to the tree before the Christmas holiday, they don't necessarily relate to each other or the tree's presentation as a whole.

development. The Growth Diagnostic approach, proposed by Hausmann, Rodrik, and Velasco (HRV), provides such a framework” (Anderson et al., 2010: 2). MCC economist Ben Dennis proposed the following in a memo published on August 22, 2006:

One of the ways in which Economic Analysis staff can better serve their transaction teams is to provide earlier guidance on both the types of projects for which financing should be considered and (later) the economic viability of proposed projects. Early guidance can avoid the waste of time and resources resulting from: (i) poor program design, and (ii) due diligence on projects that can be known to be untenable from the start (Dennis, 2006: 1).

The economists called this diagnostic approach to compact development “Enhanced Engagement”. They clearly felt that there needed to be an upfront framing of issues when MCC engaged a country. They argued to MCC senior officials that enhanced engagement through a constraints analysis would save transaction teams time and effort and improve proposal and compact quality by focusing the transaction team, core team, and the compact development process on growth.

While others at MCC were also interested and concerned about issues of growth, senior officials were primarily concerned with operational issues. They were seeking a way to narrow the list of projects countries initially proposed to put “a little more rigor around the project identification process” (Interviewee 2J, personal communication, March 15, 2016). Narrowing down the list of potential projects was a broader MCC organizational problem that demanded a solution. One MCC economist recounted what an MCC Vice President at the time said of constraints analyses (CA): “Yes, we really need something like this, the constraints analysis can be very useful, it can help us get the dogs off the table” (Interviewee 2I, personal communication, March 4, 2016). The economist suggested that the CA “gives MCC some rationale, it gives the *appearance* of objectivity, if you will. It’s a structured approach” (Interviewee 2I, personal communication, March 4, 2016, emphasis added).

The growth diagnostics' creators claimed that through a Bayesian, data-driven process analysts could identify the most binding constraints to growth. In early documents discussing the constraints analysis, MCC economists alluded to the CA's ability to narrow priorities because they "[p]rovide guidance on the *key* constraints facing growth as soon as possible once eligibility has been determined so as to shape the proposal" (Dennis, 2006: 1, emphasis added). This would solve an operational problem for MCC by providing what one senior official called a sieve through which to funnel projects. One senior technical specialist suggested that the CA took a huge weight off of management's shoulders:

I'd say that pretty much there was a big sigh of relief that we had this analytical tool so we weren't constantly engaging in political battles with our counterparts. What I mean by that is they would put on the table projects A, B, C, and D and project B was something that no else would fund and project A was something that a very powerful minister and his cronies in the industry would want and you have to push back at that kind of politicization of choices. In many ways the CA was a pure enough methodology so that it covered their asses as well as ours for the choices being made (Interviewee 2G, personal communication, February 26, 2016).

MCC economists' appeals to Harvard professors' academic authority provided a source of legitimacy for constraints analysis work at MCC. MCC economists developed links with growth diagnostics' creators at the Kennedy School. Several MCC economists sought their counsel on the implementation of growth diagnostics at MCC and Ricardo Hausmann and MCC economists communicated regularly (Interviewee 2C, personal communication, February 18, 2016). MCC economists presented their work at a Kennedy School International Development Seminar and Ricardo Hausmann presented on growth diagnostics at MCC headquarters. This legitimacy was sufficient for economists to gain senior officials' consent.

### **Building a Diagnosis**

Before we can understand how constraints analyses conflicted with other expert groups' planning approaches and rationales we must understand how economists diagnose countries'

constraints to growth. There were three main, multifaceted steps to completing a constraints analysis: data collection, conducting tests, and report writing. These steps involved identifying the CA team, desk research, work-planning sessions to organize data collection, running the tests on assembled data, discussing results, and writing the report. This all required a considerable amount of human and technical resources that needed to be marshaled and maintained over several months.

**Human Resources.** The CA was an economist-led process. While early CAs left room for business leaders to shepherd the CA process, later MCC guidance emphasized the role of economists, “MCC expects that core teams will be staffed with an economist as early as is practical who will serve as the CA team leader” (MCC, 2013: 3). For example, for El Salvador’s second compact Carlos Acevedo, President of El Salvador’s Central Reserve Bank, led the Salvadoran CA team. In Malawi, Alex Gomani, an economist at the Ministry of Finance and former deputy chief economist to President Hastings Banda, led the CA. While there were exceptions to this, economist-led CA teams were the norm and notable local economists working in either government or academia carried out the technical work.

MCC was one of, if not the, first development organizations to systematically incorporate growth diagnostics as a planning tool. This was a US economic method that development agencies were quickly attempting to make a global standard. The overall framework was foreign even to many economists and especially those in MCC partner countries. When MCC economists realized that partner countries’ core teams could not conduct CAs on their own, they employed recent MPA/ID graduates as consultants (Interviewee 2I, personal communication, March 4, 2016). While these consultants were helpful in introducing the decision tree framework, gathering data sources, and keeping the process on schedule, additional intellectual

resources were still necessary for some of the analysis, interpretation, and writing. For this, MCC drew on institutions with a long history of economic analysis—multilateral development banks. Senior economists from the World, Asian, African, and other development banks often partnered with in-country core team members to conduct CAs and write CA reports.

Despite the involvement of consultants and senior economists from other agencies, there were still usually anywhere from half a dozen to a dozen core team technical members involved. For example, Zambia had nine members and Liberia had eleven (Ministry of Finance and National Planning, 2010; National Millennium Compact Development Project, 2013). While the Kennedy School and development bank consultants were doing the high-level framing and analytical work, local economists, along with their local sector colleagues, often served as data providers. They carried out analyses too, but often in silos for specific diagnostic tree branches while the consultant or advisors steered the overall process. These local economists were nonetheless a key part of the CA network. They liaised with ministry staff, private sector and civil society groups, and academics, among others, to ensure that data were made available to the consultants and advisors. As an MCC economist put it, “The CA was not a desk exercise, you could not do a CA from DC alone. You needed the in-country data and experience to complete the CA” (Interviewee 2F, personal communication, February 23, 2016).

**Data (Re)sources.** The CA is not explicitly quantitatively data-driven, but numbers are at its heart. There is room and flexibility for qualitative analysis and narrative to enter the CA. Other economic reports such as PRSPs, relevant growth diagnostics, and donor-supported sector reviews are consulted and focus group discussions with targeted populations or places are also sometimes held. In the end, however, the CA’s promoters believe that the four diagnostic tests Hausmann and his colleagues recommend in the *Mindbook*, supported by benchmarking,

surveys, and other data-driven approaches, “elevates the discussion above the usual discourse of opinions that are often more informed by narrow interests than empirical evidence” (MCC Department of Policy & Evaluation, 2010: 4).

The CA includes not just economists and others conducting the analysis but the myriad data sources that form the foundation of its results. Broadly speaking, the CA requires information on levels, trends, and cross-country comparisons with respect to a variety of variables and parameters, on both the micro- and macroeconomic levels, as well as qualitative evidence indicating the presence of constraints” (MCC, 2009: 4). Suggested macroeconomic variables included aggregate public and private investment, factor prices such as wages and interest rates, domestic and foreign savings levels and rates, inflation, fiscal balance, public debt, the trade balance, and the current account. Suggested microeconomic variables included educational attainment levels, health of the labor force, borrowing and lending flows, quality of economic and political governance, the “cost of doing business”, and quantity and quality of infrastructure (MCC, 2009). MCC guidance on conducting CAs included a 13 page Technical Annex that went into detail on these data sources for each possible constraint.

My comprehensive review of 12 CAs from as many compact countries identified and classified each analysis’ primary and secondary data source. This analysis resulted in 33 individual sources, which are outlined in Appendix C along with the percent of analyses that relied on a particular source. One unmistakable feature is the World Bank’s dominant role. The World Bank is the primary data source for 33 percent of all analyses in CAs.<sup>33</sup> This figure does not include some investment climate diagnostics that are funded by the World Bank such as the

---

<sup>33</sup> The World Bank relies heavily on local data sources to compile its databases and indicators. Thus, distinguishing how “national” a “World Bank source” is can be impossible and perhaps unnecessary



Africa Infrastructure Country Diagnostic. National government sources also play a dominant role contributing to about a quarter (23%) of all analyses. This jumps to close to a third (31%) if national central banks are included. The United Nations, despite its important global role, plays a marginal role at seven percent. Two sources on “economic competitiveness”—the World Economic Forum and Economist Intelligence Unit—contribute almost as much as the whole UN system at six percent. Scholars and think tanks make up less than four percent.

It is more informative to understand the sources used for specific branches of the diagnostic tree (see Appendix C). This is accomplished by linking data used for individual analyses to the branches, or constraints, they were meant to interrogate. One major data source is the International Monetary Fund and national central banks. This is unsurprising given these institutions’ historic role in macroeconomic and financial management. They are the primary data source for possible constraints under the high cost of finance branch and for the constraint of macro risks. Another data source that stands out, perhaps less obviously, is competitiveness surveys. These include “enterprise”, “doing business”, and “executive opinion” surveys conducted primarily by the World Bank and World Economic Forum. Competitiveness surveys interview a range of small, medium, and large formal enterprises as well as foreign and multinational corporations operating in a country. They are a dominant data source for analyzing micro risks, infrastructure, and information externalities as possible constraints to growth.

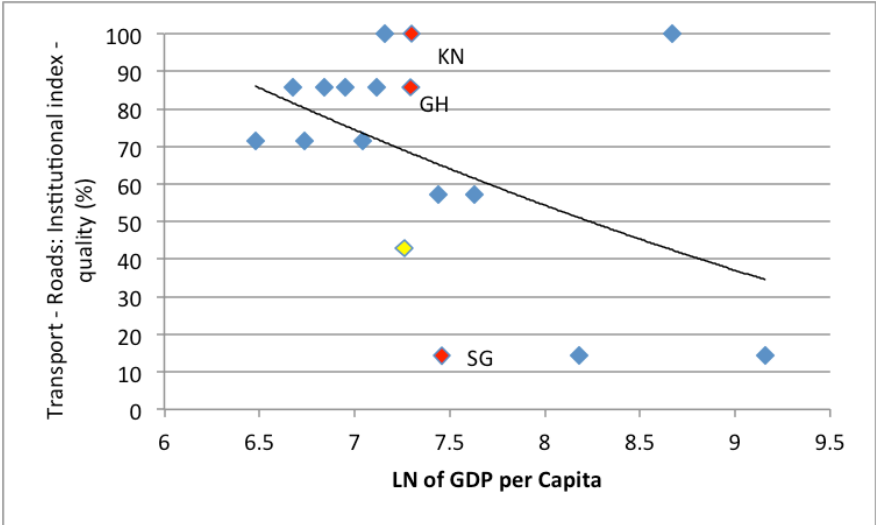
**Running Diagnostics.** For each compact in development, the CA team analyzed this data and conducted diagnostic tests on plausible constraints to growth. The CA used a variety of diagnostic tools, but MCC recommends one in particular: benchmarking. Benchmarking a country against “similarly situated comparison countries” is part and parcel of the CA: “By assessing the country of interest against plausible comparator countries...we may identify

constraints to growth causing the country to lag behind its potential growth path” (MCC, 2013: 5). Benchmarking was used nearly half the time (47%) for all analyses in the CAs examined. Most of these were snapshots of cross-country comparisons at a point in time but time-series analyses were also frequently employed when data was available. For a detailed assessment of analyses used in MCC’s constraints analyses see Appendix C, Table C-6.

The CA team usually conducted four tests to determine which of the ten possible constraints were binding, or a primary bottleneck to growing private investment. Test 1 seeks to understand if the price or shadow price of a constraint is high. When prices appropriate to the constraint are available, such as interest rates, they are used. When such prices are unavailable, the analyst employed shadow prices. A shadow price substitutes for unknowable or difficult to calculate costs as when “perfectly competitive markets” are distorted and market prices are unavailable or give misleading signals. Dreze and Stern (1990) define them as “the social opportunity costs of the resources used” (Drèze & Stern, 1990). MCC defined them as “the marginal value to the entire economy of the opportunity cost of an additional unit of the factor” (Partnership for Growth, 2011). Economists determined whether a shadow price was high using proxy prices such as gross domestic product per capita, returns to an investment or public good (e.g., an extra year of schooling), or costs to economic activity when the constraint is present (e.g., economic losses due to power outages). A bivariate analysis comparing the proxy price to the presence of the constraint or its removal across several countries usually forms the basis of this test (see **Figure 5.5**). The further a country’s data point falls below the trend line, the more potentially binding is the constraint. For example, the diamonds labeled “SG” and “GH” represent the relationship between road quality and gross domestic product (GDP) per capita in

Senegal and Ghana, respectively. Economists interpret road quality as thus presenting a greater bottleneck to growth in Senegal than Ghana.

**Figure 5.5.** Example of Test 1: Shadow Prices for Transport as a Binding Constraint



Source: MCA-Benin. (2012b). 4: Test analyses for Roads. Contonou: Millennium Challenge Account-Benin.

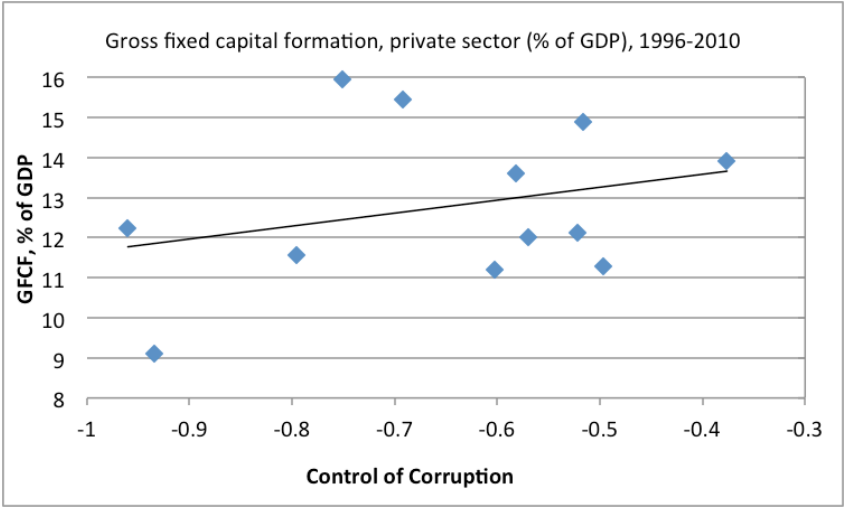
Notes: The Natural Log of GDP per Capita is a Shadow Price for Road Quality. Diamonds represent country-level data.

Test 2 examines how the relief of a constraint relates to investment or growth levels.

This is referred to as an economic actor’s “impulse response”. If a constraint is binding, growth or investment should increase as it is relaxed. For example, an analyst may ask if a denser road network correlates positively with higher growth or investment levels or if improved transparency or lower corruption have a positive association with investment levels. Bivariate analyses comparing investment or growth levels with the intensity of the constraint across several countries form the basis of the test (see **Figure 5.6**). The further a country’s data point falls below the trend line, the more potentially binding the constraint in that country. For example, a diamond-shaped data point represents a country’s relationship between its gross fixed

capital formation (the aggregate value of physical fixed assets minus land) and a composite indicator of corruption from the World Bank. If a country's data point falls below the trend line corruption may be hindering capital formation in that country relative to its peers.

**Figure 5.6.** Example of Test 2: The Relationship Between Investment Levels and Corruption



Source: MCA-Benin (2012a). *Corruption – Test 2*. Conotonou: Millennium Challenge Account-Benin.  
 Note: GFCF stands for Gross Fixed Capital Formation. Diamonds represent country-level data.

Test 3 examines how private actors try to circumvent the constraint. The more binding the constraint is the more intensive the circumventing activity will be. Examples of private actors' coping behavior include firms training their workers in skills that schools do not provide, using generators to supply electricity, paying extra to use air transport rather than road or rail, and paying bribes (see **Figure 5.7**). Analysts try to identify coping mechanisms, quantify their frequency, and aggregate costs to firms and other actors. For example, economists might examine competitiveness surveys to assess the percent of firms that employ private generators for electricity and then try to monetize the cost of running those generators.

**Figure 5.7.** Test 3: Examples of Coping Mechanisms

<b>Private responses to constraints</b>	
<b>If this constraint is binding...</b>	<b>then we might expect to see:</b>
Labor market regulations	Higher than normal levels of informal employment
Contract enforcement	The emergence of extra-legal contract enforcement mechanisms like trading within social groups or organized mafias enforcing contracts (Dixit 2005)
Electricity infrastructure	Many businesses investing in their own generators
Crime and security	Large outlays for private security guards
Uncertainty of monetary stability	Dollarization, use of inflation indexed contracts.
Coordination failures in the discovery of new activities	Vertical integration in new successful business. Efforts for industry groups to share costs in feasibility studies for new sectors and markets
Low appropriability due to high taxes	Greater use of cash for business transactions

*Source:* Hausmann, R., Klinger, B., & Wagner, R. (2008). *Doing Growth Diagnostics in Practice: A “Mindbook”* (CID Working Papers No. 177). Center for International Development at Harvard University.

Test 4 focuses on which sectors or economic activities thrive in a country. The less consequential a constraint is to an economic activity, the more likely it is to survive. This is referred to as the “camels and hippos test” (see **Figure 5.8**). The authors invoke the metaphor of camels and hippos to demonstrate that animals that require less of a resource, in this case water, survive and thrive in the Sahara while those more reliant on the resource, such as hippos, are absent.<sup>34</sup> For example, in countries with arid environments where water is a constraint, water-intensive industries such as textiles should be rare. This may not, however, affect the presence of call centers whose most important input is a strong and reliable telecommunications network. This fourth test is infrequently conducted and the one that faces the greatest data availability and measurement challenges.

---

<sup>34</sup> The authors, who speak a great deal about innovation in the economy, completely ignore the role of evolutionary environmental adaptation in allowing camels to thrive in the desert, but this contradiction will not be addressed here.

**Figure 5.8.** Test 4: A Search for "Camels and Hippos"

**Examples: Looking for camels and hippos**

Finding Camels and Hippos	
If this constraint is binding...	then we might expect to find:
Labor market regulations	Small firms that can operate informally survive as do large firms that can utilize economies of scale. Missing middle.
Contract enforcement	Firms that can vertically integrate or access extra-legal enforcement.
Infrastructure	Few businesses heavily dependent on infrastructure. For example: India has more electricity camels and China more electricity hippos.
Finance	Larger firms with the ability to self finance succeed. International firms with the ability to borrow abroad are prevalent.

Source: Hausmann, et al. 2008

28

*Source: MCC (n.d.c). Introduction to Constraints Analysis (a.k.a. Growth Diagnostics). Millennium Challenge Corporation. Washington, DC.*

According to an MCC economist, tests 1 and 2 are the most consequential. If these tests do not show a constraint to be a problem or binding, the constraint is likely discarded (Interviewee 2M, personal communication, March 22, 2016). Tests 3 and 4 are weaker and result in more ambiguous conclusions. The results may not automatically exclude a constraint and more of a discussion and negotiation can ensue about their inclusion in a CA report. The data analyzed and the people doing the analysis also matter. According to the same economist, “The quality of the test depends on whether data is convincing or all the alternative hypotheses have been considered” (Interviewee 2M, personal communication, March 22, 2016).

### **Conflicting Rationalities in Compact Design**

Growth diagnostics are congruent with the so-called Finance Ministry tendency, or rationale, held by the Finance Ministry community in development. At MCC, the economics team was most closely aligned with the Finance Ministry tendency. Growth diagnostics contended with the so-called Civil Society tendency, or rationale, held by the Civil Society

group. At MCC, the environmental and social assessment (ESA) team was most closely aligned with the civil society tendency. The main points of contention between the two groups were over whether and how to address inequality. The faith of MCC economists in competitive markets as a solution to growth made them more tolerant of inequality. They believed that competitive markets and market-led development could resolve structural inequalities over the medium to long term. The civil society group believed that structural inequalities affecting social institutions also affected markets: markets cannot resolve inequalities on their own, and marginalized groups needed to be purposefully targeted. National-level sectoral reform programs tailored to aggregate private investment would not suffice.

**The Finance Ministry Rationale.** During consultations for the 2000 World Bank World Development Report on “Attacking Poverty”, the director of the consultative process, Ravi Kanbur, found that there was considerable disagreement between the Finance Ministry group and Civil Society group on the nature of economic and poverty reduction policies and programs and how their consequences should be assessed (Kanbur, 2001). The Finance Ministry group included, broadly speaking, members of finance ministries in the Global North and South, economic policy managers at the IFIs, economic analysts, and orthodox academic economists. The Civil Society group included, again, broadly speaking, analysts and advocates in a broad range of advocacy and NGO groups in the Global North and South, members of certain United Nations specialized agencies, social sector ministries in the Global South, and many academic non-economists (Kanbur, 2001).

According to Kanbur, members holding one rationale assailed the motives and, importantly, analytical approaches of those holding the opposite rationale. When it came to issues of economic policies and their effects, potent disagreements remained and the two groups

possessed, as Kanbur described it, “differences in perspective and framework” (Kanbur, 2001). The disagreements centered on how economic policies’ effects on distribution and poverty were assessed. These differences in “aggregation, time horizon, and market structure” can be distilled as differences in how the two groups approached inequality. The main contention was whether, as the Finance Ministry group believed, inequalities between people and places within a country would be worked out naturally over time as constraints to “mutually beneficial exchange” in national markets were relaxed, or, as the Civil Society group believed, policies and programs should be designed to address *specific constraints* faced by historically or spatially marginalized people that may not rely on market exchange because mainstream economists’ market assumptions were unrealistic abstractions (Kanbur, 2001). While Kanbur’s analysis was focused on the structural macroeconomic, financial, and trade policies of the 1990s, much of the Finance Ministry tendency’s features remain salient to contemporary development approaches including MCC’s constraints analyses. In modified form, these disagreements spilled over into the MCC.

**The Finance Ministry Tendency and Growth Diagnostics.** The Finance Ministry group’s faith in competitive markets as an institutional solution to growth makes its members more tolerant of inequality. The group is influenced by orthodox economic thought and believes that maximizing mutually beneficial transactions is the best or most efficient way to address inequality and issues of distribution. Its members believe that when markets are free of failures market transactions are mutually beneficial and those transacting do not have power over each other (Kanbur, 2002). When markets have failures, such as the under-provision of a public good like education that make transactions inefficient or asymmetrical information that might give one actor power over another, the government should actively address such failures. These interventions should be aimed at markets and not inequality, however, because issues of



efficiency and equity should be addressed separately (Kanbur, 2002). Adherents of the Finance Ministry tendency design and evaluate policies and programs in national aggregative terms (Kanbur, 2001). They are less interested in whether a set of policies or programs benefits, for example, the poor or rural areas relatively more than the rich or urban areas so long as national average per capita gross domestic product (GDP) increases. They evaluate the impact of economic policy reforms with national level statistics and indicators on GDP and poverty. Finally, the Finance Ministry group interprets the value of interventions over a “medium-term” time horizon of five to ten years (Kanbur, 2001). Economic reforms may generate inequality at first but through free and competitive markets and increased aggregate growth, inequality will be reduced over time. Equilibrium or equality along some metric such as income between groups and places will be achieved as firms and other economic actors seek out abundant inputs with lower marginal costs (Kanbur, 2001). Competitive markets and market-led development can resolve structural and enduring inequalities via the price mechanism.

Growth diagnostics’ creators shared the Finance Ministry’s rationale. They believe that a Pareto optimal allocation of resources that maximizes welfare through competitive markets is limited by a host of “real world complications” rooted in “market imperfections” (Headey, 2009; Rodrik, 2009). Economists’ responsibility is to make an effort to address these market failures to bring out their welfare maximizing potential.<sup>35</sup> They assume that mutually beneficial market transactions arise because individual agents engage in “purposeful behavior” to make themselves better off (Rodrik, 2009). Growth diagnostics share fealty to an evaluation of interventions at the national level. Ricardo Hausmann wrote, “Growth diagnostics goes from a very aggregate

---

<sup>35</sup> Their approach can be better described as “neoclassical structuralism” because it acknowledges that markets are not “perfect”: if left to their own devices markets will not produce equilibrium prices and maximize surplus because they might possess significant flaws that vary from country to country and may require state intervention to correct.

outcome, such as the growth rate of an economy, to its potential causes. As such, it is a top-down approach” (Hausmann et al., 2008: 20). They are focused on constraints that differ *across* but not necessarily *within* countries. Finally, growth diagnostics claim to get at the “policy roots” of what constrains economic growth. As such, they are focused on economic policy and institutional reform. Addressing the root causes of what ails an economy is a longer-term process that requires a longer time horizon to realize results than, as the Civil Society group might favor, the distribution of immediate benefits through subsidies, social protection programs, or basic services. In the Finance Ministry group’s view, targeting the poor in the latter way would only delay the actions necessary for self-sustaining growth and poverty reduction. In the words of one MCC economist, “The policy reform issue is important if you want to get the patient not only off the table but get them walking again” (Interviewee 2C, personal communication, February 18, 2016).

**The Finance Ministry Rationale at MCC.** MCC and in-country economists were more closely aligned with the Finance Ministry tendency. This rationale determined what data MCC economists trained in the neoclassical tradition looked at, ignored, and collected (Interviewee 2L, personal communication, March 22, 2016). MCC economists relied heavily on competitiveness surveys to conduct CAs (see Appendix C, Table C-4). These were often the only surveys of private firms’ owners and managers. They are produced by organizations that share the Finance Ministry tendency and capture the views of a relatively small sample of business owners in urban areas in the formal sector. For example, the World Bank’s Doing Business survey’s indicators refer to a specific type of business, generally a local limited liability company operating in the commercial capital (International Finance Corporation, 2011). The focus on a specific, narrow set of private investments and firms meant that households and their economic activities, which

include informal household enterprises, went ignored. As a result, the informal sector became a blind spot and the CA failed to “see” the informal sector’s various economic constraints. This bias extended to agricultural workers and women, which are populations that disproportionately make up the informal sector (Lund & Srinivas, 2000).

It is not entirely clear, however, if MCC economists would have sought out or employed data sources that focused on the informal economy even if they had been available. MCC economists saw the informal sector as a problem that needed to be corrected. As one MCC economist put it, “To do a CA one needs to have some sense of how development needs to happen...social sector folks are more concerned with work and investment in the informal sector but I’m looking 20 years into the future [and asking], ‘What’s going to create new jobs?’ ” (Interviewee 2N, personal communication, March 22, 2016). This view implies that “new jobs” were those created in the formal sector resulting from a process of long-term growth.

Virtually all the data used to measure income, growth, and investment in a MCC constraints analysis was collected at the national level. Data on GDP and investment, which play an important role in the first two tests, is almost exclusively available from the IMF, multilateral development banks, and national governments. The IMF and multilateral development banks, as multilateral institutions whose members are nation-states, have since their founding organized data collection at the national level. Countries in the Global South, in turn, are drawing on data collection infrastructure for a construct called “the economy” developed during colonialism (Mitchell, 2002). This reliance on national-level data is exacerbated by MCC economists’ emphasis on using comparator countries for benchmarking exercises that require data on similar scales across units of analysis. Thus, this comparative approach, which is a tradition in development economics, extends the logic of the competitive market to economic epistemology.

The tests MCC economists conducted as part of the CA also shared the Finance Ministry tendency. Test 1 manufactured fictitious markets characterized by perfect competition and equilibrium when no such markets existed. It achieved this by estimating shadow prices where they did not exist or imputing shadow prices to those things that are not or cannot be subject to market exchange. Test 2 examined the relationship between measures of capital formation (e.g., investment or growth levels) and the presence or absence of a constraint. Where the constraint was present and capital formation was low the constraint was thought to be binding. Such relationships were usually examined at the national level without regard for the intranational distribution of capital. If constraints must be relaxed in a way to maximize capital formation, investments were usually made in areas with abundant complementary factors such as skilled labor or where economies of scale already existed such as large cities.

The Finance Ministry tendency informed the way MCC economists approached a very common investment category at MCC—infrastructure. They believed that if you improve investment opportunities for formal businesses by lowering the costs of their social overhead capital, economic benefits would reach the poor and informal workers over time in the form of increased jobs. For example, Test 1 analyzes the operating revenue losses from interruptions in infrastructure services such as water or electricity. When the impacts of these stoppages, such as blackouts, are estimated for different groups of people, those that are already better off are estimated to suffer disproportionately more. When aggregated, such shadow prices signal that investments to ameliorate stoppages will reduce operating losses and raise revenues far more for large, formal firms than informal household activities. One country director involved in designing a compact explained it this way:

If you got blackouts in the southern part of [Country], they're high valued blackouts, because you've got businesses and so forth that depend on the electricity. So if you are a

bottler and lose electricity even for a few minutes it could cost hundreds of thousands of dollars. Even though there are lots of benefits in terms of residential usage like education or chores and things like that, those are more difficult to value so we don't value them...that cost of non-served kilowatt-hour is a lot less (Interviewee 2R, personal communication, March 30, 2016).

Finally, MCC economists prioritize private-sector reform of infrastructure sector institutions over an infrastructure sector's storage capacity or service expansion, which affects the time horizon they use to evaluate outcomes. MCC economists made it clear early on that they preferred policy and institutional reform to service expansion: "For some sectors, such as energy, under-capacity and lack of growth may be due more to a deficient regulatory environment (e.g., tariffs set too low) that stifles private investment and prevents sufficient cost recovery and maintenance...it should be clearly communicated to the recipient country that funding for these kinds of projects is unlikely unless certain reforms are enacted" (Dennis, 2006: 2).

**The Civil Society Rationale.** The Civil Society group believed that capital and labor were not mere factors of production but a set of social and power relations; the aggregate national picture hid regional differences and pockets of distress; and policies should be assessed according to their short-term consequences (Kanbur, 2001). This is because the Civil Society tendency takes a political-economic and broader social, historical, and cultural approach to development. The Civil Society group views market structure as riddled with market power. Market transactions usually benefit one party at the expense of another. Capital and labor are not simply factors of production but a set of social relations where actors belonging to each group are engaged in constant negotiation (Kanbur, 2001). Relationships between them are often zero-sum and market liberalization can give capital the upper hand. Also, local markets and economies are segmented, which makes problematic neoclassical economic theory about self-

interested maximizing agents and equilibrium prices reducing special or demographic inequalities over time. Marginalized groups do not necessarily receive the opportunities or pay of their non-marginalized peers and inequalities in existing social overhead capital such as infrastructure limit how much core and peripheral regions can be integrated (Kanbur, 2001).

The Civil Society group's concern for inequality leads them to consider development interventions and their outcomes at a lower level of aggregation than the nation-state (Kanbur, 2001). Its members are more interested in whether policies or programs benefit those worse off relatively more than the better-off regardless of whether everyone benefits in absolute terms. They argue that local conditions may differ significantly from the picture aggregate, national-level economic and poverty statistics paint and may obscure pockets of distress (Kanbur, 2001). They are committed to and actively engaged in addressing conditions in these distressed areas or among distressed groups.

Finally, the Civil Society tendency has a much shorter time-horizon than five to ten years. Those working directly with the poor and vulnerable are focused on the short-term consequences of economic reform or immediate benefits from development projects (Kanbur, 2001). They do not share the Finance Ministry group's faith in competitive market equilibrium. This leads the Civil Society group to have a greater tolerance for "pork barrel" spending. Some projects, while economically inefficient in the Finance Ministry group's eyes, can be justified on grounds of distributional equity (Ellwood & Patashnik, 1993).

**The Civil Society Rationale at MCC.** The Civil Society group was represented by set of actors dispersed across MCC's various technical functions and in-country core teams. Members of MCC's Environmental and Social Assessment team were some of the strongest advocates of the Civil Society tendency. They argued that the CA methodology did not adequately account

for segmented economies and labor markets in the Global South. One ESA specialist asserted that the CA exhibited a “formal sector bias” in its use of data sources and surveys to examine constraints (Interviewee 2L, personal communication, March 22, 2016). This in turn had implications for how CAs addressed issues of poverty and gender.

A review of early CAs by MCC’s environmental assessment team critiqued the CA’s “rather stratospheric trickle-down look” into countries’ situations. As the author put it, “A limitation of the CA is the national level approach which precludes consideration of existing documentation of 20 identified pockets of poverty in the country” (Feld, 2009: 3). The ESA specialist’s principal point was that disregarding important in-country variation precludes bottom-up approaches to growth and development rooted in sub-national realities (Feld, 2009). Investment should not be directed on the basis of shadow-prices, which assume or imagine a competitive market, or relationships to capital formation without regard for where capital is specifically invested. Rather, they should be channeled to marginalized populations such as informal workers, women, and the rural poor; sectors in which lower-income individuals are disproportionately employed such as agriculture; and marginalized regions in the periphery. These groups, regions, and sectors often overlap and provide a clear picture of where investment is most needed. Because these populations and places are often relatively lower-income and face any number of multidimensional hardships, the Civil Society group does not view investments in them as “transfers” or “wasteful” if they do not generate long-term, aggregate growth. As long as investments address these inequalities they are worth making.

Investments in policy reform rarely meet the Civil Society tendency’s criteria. The Civil Society group would, for example, prefer increasing immediate access to basic services rather

than reform of the legal, policy, and administrative institutions responsible for delivering those services.

Most early MCC compacts and their respective proposals implicitly advanced the Civil Society tendency by emphasizing intranational differences and targeting the majority of funds to a specific region or zone of the country. These regions were often those that were disproportionately poor or rural. Economic geographers would deem that these proposals and the eventual compacts exhibited a “place-based approach”. Such an approach considers that geographical context, as defined by the economic, social, cultural, and institutional dimensions of a locality, really matters (McCann & Rodríguez-Pose, 2011). A senior official in MCC’s development policy team compared early compacts to an approach that was popular in multilateral development banks in the 1970s when their focus was on poverty reduction, “Many of the compact proposals from our partner countries include elements that are similar to aspects of the integrated rural development approach... Integrated Rural Development Programs (IRDPs) were designed to increase incomes and other aspects of welfare among the rural poor” (Kelly, 2007: 1).

Compacts with a place-based emphasis included those with Nicaragua, Ghana, El Salvador, Mozambique, Morocco, Senegal, Mali and Madagascar. Nicaragua’s compact focused on the Northern Pacific Region, Ghana’s on the Afram Plains and Volta regions, El Salvador’s on the “Northern Zone”, Mozambique’s in the “Northern Triangle”, Morocco’s in the “Corridor of Poverty”, Senegal’s in the Senegal River Valley region, Mali’s in the Alatorra region, and Madagascar’s in the Vakinankaratra-Mania and Menabe “zones”. Nicaragua’s and El Salvador’s compacts were focused on regions disproportionately suffering from high poverty. Senegal and Mali’s compacts were focused on highly fertile areas adjacent to major rivers with promise for



agricultural development. In Morocco's MCC proposal the central government was concerned about a balance between urban and rural constituencies. One agriculture official involved in the early stages of compact development for Morocco said, "Their king wanted to show that the farmers were included in the process, he wanted to show people in the mountains that not everything was going to be absorbed by people in the city" (Interviewee 2Q, personal communication, March 29, 2016).

Even in compacts that did not have an explicit zonal focus, there was still an emphasis on populations in what Gunnar Myrdal called "the periphery" (Myrdal, 1971). Examples of this included compacts in Georgia, Armenia, and Tanzania. Georgia's compact focused on infrastructure and urban development in small municipalities outside of the capital region. Armenia's compact focused on irrigation development that was scattered but nonetheless in the poorer parts of the north and west of the country. Tanzania's compact had energy, water and road projects far from Dar-es-Salaam near the borders with Burundi and Kenya. Meanwhile, MCC's chief economist, Franck Wiebe, believed that programs aimed at "connecting or opening regions that remain on the country's periphery" were rooted in national political objectives and not economic rationality (Rose & Wiebe, 2015b). This is because orthodox economics consider such policies to be either politically motivated or misguided attempts at improving welfare.

### **Power Relations**

Despite their conflict with the Civil Society rationale, constraints analyses were incorporated into compact design with little to no resistance. This was because constraints analyses received senior officials' blessing and the ESA team's work was mostly about reducing harm (e.g., environmental permitting, proper resettlement) rather than proactively addressing inequality. Even if the ESA team took a proactive role, its members were scattered across the

agency, which made it more difficult to build a network at MCC and with like-minded social ministries in partner countries to resist the push for constraints analyses. Ultimately, MCC's adoption of constraints analysis elevated economists in the compact development process. Sector specialists were demoted and relegated to the role held earlier by economists—data providers.

**The Smooth Stabilization of Constraints Analysis.** Senior officials, to whom economic, ESA, and sector specialists reported, were comfortable handing the reins of compact development over to economists. This was because they believed economists were helping the organization make decisions based on evidence. One senior official noted, “I think [the constraints analysis] certainly increased the economists’ role early on in the process and therefore diminished the role of country programs, but that was a good thing because the economists were doing the analytical work and trying to get at a thoughtful evidence based decision” (Interviewee 2P, personal communication, March 19, 2016). Senior officials stood back and let various experts do the work. This hands-off attitude allowed the economists’ rationality to overpower other experts’ rationalities. Because identifying constraints was no easy task this also weakened senior officials’ influence over compact design and made them dependent on economists: “To actually do the constraints analysis the way [Hausmann, Rodrik, and Velasco] designed it, it was hard. You needed a PhD economist to do it” (Interviewee 2J, personal communication, March 15, 2016). Nonetheless, transferring leadership over to economists was a sacrifice senior officials felt was worthwhile.

The environmental and social assessment experts’ ability to mount a challenge to the economists and the CA was limited. Their work at the time was largely defensive—when CAs were introduced in the late 2000s ESA’s primary role was to ensure that MCC investments “do

no harm” (Interviewee 2G, personal communication, February 26, 2016). Their expertise was drawn upon after compacts had been negotiated to assess whether proposed projects posed a threat to vulnerable ecosystems, non-human species, and human populations. Their work centered on environmental and social impact assessments and resettlement plans. Moreover, ESA’s organizational position distanced the group’s members from compact development decisions.

The Civil Society group was not unified organizationally at MCC headquarters. This was important because unlike other donors and development finance agencies that have semipermanent offices in countries, most of the decision-making at MCC during compact development occurred in Washington, DC. In 2008, by which time constraints analyses were fully operational, ESA experts were split between the Departments of Compact Development and Implementation (economists were located in both departments but also the influential Department of Policy and International Relations). ESA members most sympathetic to the Civil Society tendency were located in the Department of Compact Implementation, which left them little to no influence over how compacts were designed. Their functions were also meant to manage project risks, putting them in a position to critique CA outcomes without necessarily presenting a cohesive alternative rationale for framing compact development or selecting projects. Without a network or cohesive theoretical alternative, the Civil Society tendency did not present an obstacle to MCC economists and the CA.

Ministries of Rural Development, Labor, Health, or Social Welfare along with civil society and poverty observatory groups made up the in-country Civil Society group. Their proposed projects had more “social” objectives that lacked “clear links to economic growth” and were discounted more often than Ministries of Finance, Infrastructure, or Trade and

Development (Anderson et al., 2010). The CA shaped public consultations to focus on economic growth rather than social programs and laid the foundation for possible dismissal of proposed projects that could not make a compelling case for economic growth (Anderson et al., 2010).

**The Ascendancy of the Economist.** Timing was a key factor for influencing what compacts looked like. CAs reversed the timing of when sectors and economists became involved in compact development. Prior to the CA, MCC senior officials and sometimes sector specialists would first meet with core team counterparts and team leaders and senior officials told economists to withhold judgment until enough details were available to determine a project's viability. Meanwhile, "steam would build up" around projects during negotiations and generate important political capital. One MCC economist recounted that "a lot of effort and money goes in from MCC's and the country's side in a project. Interests are ginned up in a country and momentum and interests and political commitment builds. If the project is not ultimately viable, that buildup of interests weakens the economists' decision-making power on the project" (Interviewee 2F, personal communication, February 23, 2016). According to one senior manager, "As far as economists, back in the early compacts, they were just doing some [economic rates of return analyses]" (Interviewee 2H, personal communication, March 2, 2016).

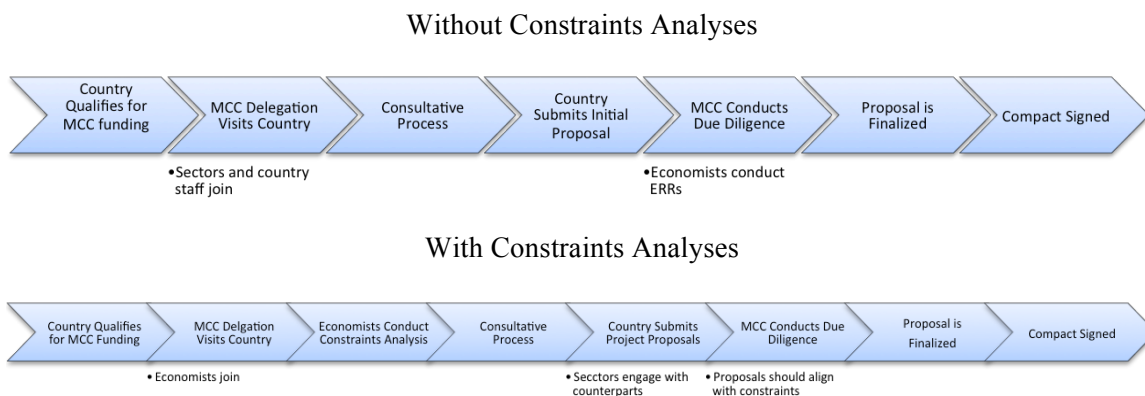
The constraints analysis changed all this (see **Figure 5.9**). After the CA, MCC economists would visit their core team economic counterparts as the first step in compact development. This was appealing to those in the agency looking to make compact development more rational:

[the CA] brings economic thinking and analytical capacity to the start of the conversation with the country, which changes the whole tenor of how you engage with a country. If you start by showing up with your ag[riculture] people and your infra[structure] people and your country relations people, versus if you show up with economists, and you say to the country give us your economists, that's who is having the initial set of conversations ...for somebody who is interested in getting evidence into the policy process that's a

fantastic place to start and it sets the tone for a relationship that I think carries forward (Interviewee 3C, personal communication, May 26, 2016).

According to an MCC economist, the CA “helped move economists concerns up front—if you can get projects designed and approved that are consistent with the constraint you are halfway there” (Interviewee 2F, personal communication, February 23, 2016). A sector specialist dubbed this shift in roles and responsibilities as the “ascendancy of the economist” (Interviewee 2E, personal communication, February 24, 2016).

**Figure 5.9.** Compact Development Process Charts



Source: Author.

If the economists were ascendant, then the sector specialists were descendant. After the CA was instituted, sector specialists “don’t go out to the countries anymore to negotiate compacts” (Interviewee 2H, personal communication, March 2, 2016). Their role was pushed to the end of the compact development process similar to when and how economists would engage prior to the stabilization of CAs. This was welcome news to the economists who believed that “[s]ector folks are not the most objective and may not be the best to send out to the country...If you’re a carpenter, everything looks like a nail” (Interviewee 2I, personal communication, March 4, 2016). The sector specialists’ role was relegated to what economists mainly did before CAs—

identifying and providing relevant data sources. Sector specialists in the partner country provided data sources relevant to their sector. More often than not, MCC sector specialists would be brought in after the CA was completed to provide data and analysis for the sector to identify and vet projects (Interviewee 2M, personal communication, March 22, 2016).

Nonetheless, the economists maintained editorial control of the CA findings and the process in general. Sector specialists went along with this subordinate role because contributing data and input was a way to have some say in compact design.

### **The Power of Method: Diagnosis**

The power of constraints analyses affected not only power relations at MCC but organizational outcomes in the form of the diagnoses that could be reached. By 2015, MCC constraints analyses informed fourteen compacts. The economic analysis team identified, on average, three binding or highest priority constraints to growth. While MCC economists decried the subjective biases of sector specialists, bias had now shifted to the decisions that economists made in diagnostic analyses such as which constraints to consider and test. Constraints regarding infrastructure, micro risks to private investment, and human capital were tested most often. Data availability and quality were often themselves a constraint on what could be reasonably examined.

**Diagnoses.** In addition to narrowing the focus of compact investments, the other motivation for adopting CAs was the elimination of so-called selection bias. Economists believed that sectoral preferences at MCC and ministers' proclivity for pet projects led to a politicized compact negotiation process where individual interests would trump objective, rational decision-making. This, in turn, would hurt prospects for growth and the poor. The laborious, analytical, data-driven CA process did not, however, eliminate bias—it changed the

nature of bias. Bias in sector selection became a feature of political values and assumptions in the growth diagnostic method and decisions made by MCC economists.

Once data was gathered the CA team began a process of elimination. Just as there was a laundry list of possible projects that could be proposed, there was also a laundry list of possible constraints and sub-constraints that economists could choose to analyze. This list was narrowed down in three steps. First, growth diagnostics' creators limited the universe of constraints with economic theory and existing econometric research. They named ten plausible constraints that they organized into a decision-tree (see **Figure 5.1**). Second, the analysts conducting the CA excluded constraints based on a desk review of earlier economic reports and growth studies (Interviewee 2M, personal communication, March 22, 2016). Third, analysts conducted the four diagnostic tests.

Table 5.3 demonstrates that certain constraints were examined far more than others. Close to 60 percent of all analyses were related to constraints falling under the branches of infrastructure, micro risks, and human capital. Some branches made up far less (e.g., coordination externalities).

Table 5.3

*Distribution of Analyses by Constraint in CA Final Reports*

Constraint	Number of Analyses	Percent (%)
Infrastructure	327	27
Micro Risks	212	18
Human Capital	168	14
Macro Risks	110	9
Poor Intermediation	110	9
Information Externalities	89	7
Geography	79	7
Domestic Savings	66	5
International Finance	34	3
Coordination Externalities	10	1
Total	1,205	100

*Source:* Constraints Analyses' Final Reports.

These decisions and analyses determined which of the constraints to growth would be considered binding. Among the fourteen CAs analyzed, 79 total constraints were identified. Of these, 47 were binding constraints, or considered the most important barriers to economic growth (see Table 5.4). Thus, 60 percent of all identified constraints were considered binding. The constraints identified most frequently were on the infrastructure branch, usually identifying deficits in particular sectors, and micro risks branch, usually identifying governance issues such as regulation and property rights. These were also among the constraints most often identified as binding (see Table 5.4). Human capital, or deficits in education and health, represented 15 percent of all constraints. Human capital constraints were considered binding about half the time. Those three constraint categories are responsible for close to 80 percent of all the constraints identified. Poor Geography was identified only three times and every time it was



identified as binding. Poor Intermediation, meanwhile, only showed up six times. International financial and macroeconomic issues were almost never found as constraints, binding or otherwise.<sup>36</sup>

Table 5.4

*Number of Times Constraint Identified in Final Report and Classified as Binding*

Main Branch	Constraint	Classification (N)			Classification (%)	
		Binding	Non-binding	Total	Percent Binding	As Percent of All Constraints
Low Social Returns	Bad Infrastructure	18	11	29	62	37
	Low Human Capital	5	7	12	42	15
	Poor Geography	3	0	3	100	4
Low Appropriability	Micro Risks	15	6	21	71	27
	Coordination Externalities	3	2	5	60	6
	Info Externalities	1	1	2	5	3
	Macro Risks	0	1	1	0	1
Bad local finance	Poor intermediation (Access)	2	4	6	33	8
	Low Savings	0	0	0	0	0
Bad international finance	Bad international finance	0	0	0	0	0
Total		47	32	79	60	100

*Source:* Constraints Analyses' Final Reports.

By comparing the identified constraints with the sectors eventually selected for investment we see that MCC invests in a sector recommended by the CA report three-fourths of

<sup>36</sup> A large motivation for choosing these constraints implicitly includes what MCC can reasonably address. As a bilateral development finance agency that operates on the project finance model, MCC has a limited ability to address macroeconomic constraints and much more flexibility in addressing constraints in areas such as infrastructure and human capital.

the time (74 percent) (see Table 5.5). Binding constraints are almost always identified within these sectors (one exception was the decision to invest in education in El Salvador's second compact where human capital was identified as a non-binding constraint). This means that one-fourth of the time a non-constraint makes its way into the compact. It should be noted that this was most often the case in the earliest compacts designed with a CA when MCC was still transitioning towards stabilizing the method. As time passed, the correspondence between binding constraints and sectors selected for investment grew stronger. As one MCC economist put it, "It is almost always the case that if a constraint is excluded, projects won't be developed around that constraint in the compact" (Interviewee 2M, personal communication, March 22, 2016).

Table 5.5

*Distribution of Identified Constraints Matched to Final Compact Investments*

	Total Constraints	Binding Constraints	Non-binding Constraints	Percent Binding (%)	Number of Sectors	Compact Sector 1	Compact Sector 2	Compact Sector 3
Moldova	3	2	1	67	3	Irrigation	Agricultural Development	Roads*
Philippines	4	4	0	100	4*	Community Infrastructure	Roads	Tax Revenue Administration* Procurement Administration*; (4) Land / Property Rights; (5) Renewable Energy*
Indonesia	7	5	2	71	6*	Community Infrastructure*	Health	
Jordan	3	3	0	100	1	Water Supply*		
Malawi	8	4	4	50	2*	Energy*	Land / Property Rights	
Cape Verde II	8	5	3	63	2	Water & Sanitation*		
Zambia	7	5	2	71	1	Water & Sanitation*		
Georgia II	5	2	3	40	1	Education*		
Ghana II	5	3	2	60	1	Energy*		
El Salvador II	3	2	1	67	3	Investment Promotion / Governance*	Trade*	Education**
Benin II	8	3	5	38	1	Energy*		
Liberia	6	2	4	33	2	Energy*	Roads*	
Tanzania II	7	3	4	43	1	Energy*		
Morocco II	5	4	1	80	3*	Education*	Land / Property Rights*	
<b>Total</b>	<b>79</b>	<b>47</b>	<b>32</b>	<b>-</b>	<b>-</b>			
<b>Average</b>	<b>5.6</b>	<b>3.4</b>	<b>2.3</b>	<b>60</b>	<b>2</b>			

\* Identified as a binding constraint in CA Final Report; \*\* Identified as a non-binding constraint in the Final CA Report.

Source: MCC Constraints Analysis Final Reports.

According to CA reports, energy was cited most frequently as the binding constraint to growth (eight times). But roads were not far behind (six times). Education was cited most frequently but it was binding in only four instances. Irrigation does not show up at all. Only Moldova, the first country to undergo a CA, had any irrigation or agricultural development investments in its compact and its CA report did not cite this as a constraint whatsoever; its inclusion was the function of other negotiation or argumentation.

### **The Power of Method: Reshaping Participatory Planning**

The diagnoses that MCC economists arrived at through CAs influenced MCC's consultative process. The constraints analysis substituted for the development finance community's Poverty Reduction Strategy Paper (PRSP) and reduced the length, breadth, and depth of participatory processes. PRSP consultations typically gave greater voice to the rural poor. Constraints analyses were, by contrast, more dominated by economists and dismissive of rural concerns.

**PRSP-driven Consultative Processes.** As part of the Heavily Indebted Poor Countries debt relief initiative established in 1996, the World Bank initiated a poverty reduction strategy process resulting in poverty reduction strategy papers (PRSP). An ambitious strategy like the PRSP was bound to disappoint many observers. The PRSP was plagued by claims that it was not a significant enough break from earlier development bank- and donor-led approaches. Critics argued that the process continued to be government and technocrat-led, parliaments were almost always excluded from participation, and the civil society model of participation aspired to but did not reach the level of joint decision-making (Kamruzzaman, 2009). The most penetrating critique was that it co-opted civil society's energy and ideas, thereby "depoliticizing" activist energies and rendering them ineffective for genuine change (Porter & Craig, 2004). While

participation and partnership in practice fell far short of the rhetoric, there was progress in several respects, albeit incremental, and some genuine accomplishments in specific countries. A set of case studies revealed that the PRSP process brought a broader set of actors into the policy-making process and in some instances were newly created (Piron & Evans, 2004). The process expanded and deepened existing national participatory planning processes by including large numbers of participants over an extended period of time. It also presented an opportunity to expand discursive policy possibilities (Cheru, 2006; Dewachter & Molenaers, 2012).

Conducting a broad participatory process to develop compacts and inform their design was part of MCC's legislation, which stated that "in entering into a Compact, the United States shall seek to ensure that the government of an eligible country (1) takes into account the local-level perspectives of the rural and urban poor, including women, in the eligible country; and (2) consults with private and voluntary organizations, the business community, and other donors in the eligible country" (Pub. L. 108-199, 2004). For its first 19 compacts, MCC either accepted the PRSP consultative process as a substitute for its own or modeled its consultative process for a proposed compact after the PRSP process. MCC did this because it found civil society participation in the PRSP process to be "broad" and "effective" (Lucas, 2006b). For example, in Honduras "the consultative process of the PRSP included attendance by 3500 representatives of civil society organizations in 13 cities" (MCC, 2005a: 5). Most of the participatory processes took place over years, not months or weeks, with El Salvador's being one of the lengthiest: "The Government of El Salvador...proposes use of MCA funds for a development program focused on the northern region of the country. The program is the result of an eight year consultative process led by the National Commission for Development" (MCC, n.d.b: 1). According to a Center for Global Development assessment of compact development for Tanzania's first

compact, “It is certainly appropriate to rely on a good PRS[P] to identify target sectors. And it is also okay to sign on to some of the PRS[P]’s specific programs if they too reflect consultation and an assessment of poverty impact” (Lucas, 2006b).

**Constraints Analyses-driven Consultative Processes.** Following MCC’s adoption of constraints analyses, the consultative process changed. The CA had replaced the PRSP as the basis for the participation process and a source of project proposals. The PRSP continues to be consulted by the core and MCC transaction teams, as is required by MCC legislation, but plays a diminished role. According to one senior official who had been present for consultative processes both before and after CAs, “We probably did give more space to PRSPs then. The PRSP probably substituted for some of our own internal analyses, but now that we’re doing constraints analyses, we don’t rely as much on the PRSP” (Interviewee 2P, personal communication, March 29, 2016). In fact, replacing the PRSP was one of the things motivating MCC economists to adopt CAs. They believed that reliance on the PRSP for the consultative process was too “wishy washy” (Interviewee 2O, personal communication, March 25, 2016). The CA was meant to “structure the dialogue as much as possible” and serve as a “technical resource, a set of vocabulary” to keep the conversation focused (Interviewee 2O, personal communication, March 25, 2016). As one senior official remarked, “Ultimately the constraints analysis has completely transformed the way MCC works, it has almost become a religion to refer back to the constraints analysis during discussions and ask, ‘what constraint are we relaxing?’ ” (Interviewee 2H, personal communication, March 2, 2016). A free flowing and extensive consultative process was not conducive to rapidly narrowing the priorities of the compact.

Structuring the consultative process around the CA, however, had implications for the nature of local participation and compact design. Like earlier consultative processes, consultation in later compacts involving CAs was diverse with differing degrees of success. The general differences were that the length of time for consultation was dramatically reduced, often times to just one week, as was the case for Benin's and Morocco's second compacts. Much of the time that was previously spent on consultation in compact development was now spent conducting the CA. As a result, fewer participants were consulted. The focus was generally less on the rural poor or members of the informal economy and more on members or owners of formal private-sector enterprises. Most importantly, earlier consultative processes were rooted more in dialogue while consultative processes governed by the CA were focused more on validating identified constraints. An example from Benin's second compact illustrates the new consultative process:

Consultations around the Constraints Analysis for Benin's second compact were held in each region over approximately a week, and separate consultations were held in Cotonou for sectoral input and to synthesize information received and incorporate it into the CA. The agenda at each session included: an introduction, validation and discussion of CA priorities, guided discussion on key priorities to understand how the constraints are experienced by different stakeholders, brainstorming solutions and discussion on next steps and how participants could remain engaged (MCC, n.d.a: 6).

The focus of the consultations became about how the identified constraints impacted participants. The underlying assumption was that the constraints *did* impact them; the questions were focused on how. The CA-led consultative processes diminished the length and depth of MCC's consultative process and constrained the discursive space for new kinds of ideas to arise and shape compact design.

## **The Power of Method: Implications for Compact Composition**

As intended, CAs concentrated investments in a smaller number of sectors and projects. They also, however, shifted discretionary authority away from presidents and ministers in partner countries to White House and MCC officials in Washington, DC. Energy, rather than Transport or Agriculture and Rural Development (ARD), became the dominant sector in MCC's investment portfolio in large part because of the Obama administration's Power Africa program, sometimes in contrast to the priorities expressed by country officials. Because of their neoclassical economic approach, CAs shifted investments from rural to urban areas. Finally, their focus on reform shifted investments from "hard" infrastructure to "soft" policy and institutional reform; MCC compacts began to resemble the World Bank's sectoral structural adjustment programs of the past.

**The Distribution of Investments Across Sectors.** One of the reasons senior official consented to the CA was to guide the decision-making process and narrow down the choices for investment. The CA succeeded at this. Compacts with CAs had on average half the number of sectors and 20 percent fewer activities (see Table 5.6). Compacts developed without CAs represented 18 sectors while those developed with CAs represented 13 (a 28 percent decline). The drop in total activities was more drastic. Compacts without a CA represented 212 activities and those with a CA represented 113—the number of activities was cut almost in half. The size of the compacts in terms of funding stayed about the same. The median size of a compact without a CA was \$354 million versus \$372 million with a CA. The average size of compacts without CAs was slightly higher (\$401 million versus \$371 million) but this is because Tanzania's compact was unusually large at \$787 million (in 2015 dollar terms).



Table 5.6

*Summary Statistics for Compacts Developed with and without Constraints Analyses*

		Number of Sectors	Number of Projects	Number of Activities	Compact Funding (\$)	Project funding (\$)	Activity funding (\$)
First 19 Compacts: No Constraints Analyses (Signed between 2005 and 2009)	Mean	4.3	3.3	13	401	123	32.9
	Median	4	3	14	354	104	12.7
	Min	1	1	2	76.1	8.6	0.12
	Max	7	5	19	787	413	390
	Total	18	57	212	7,218	6,978	6,978
Recent 14 Compacts: With Constraints Analyses (Signed between 2010 and 2015)	Mean	2	4.1	10.4	371	118	39
	Median	1	3	9	372	88	15.7
	Min	1	1	2	69.1	5	0.49
	Max	6	7	20	638	353	296
	Total	13	43	113	5,567	4,406	4,406

*Source:* Compacts' Summaries of Multi-Year Financial Plans.

*Notes:* All dollar figures in millions of dollars in 2015 terms.

Not only did CAs reduce the number of sectors and activities in MCC compacts they also affected *what* MCC invested in. Compacts designed with CAs invested heavily in infrastructure and human capital just as the CA reports recommended. But this does not tell the whole story. Infrastructure was always popular at MCC. However, a shift among sectors occurred. Previously, MCC invested heavily in transport and agricultural development, which included infrastructure such as irrigation. After CAs, MCC mostly invested in the energy and water and sanitation sectors (see Table 5.7).

Table 5.7

*Distribution of Sectors Among Compacts Developed with Constraints Analyses*

Rank based on funding	Sector	Funding (\$)	Percent of Total (%)
1	Energy	1,702,203,474	38.6
2	Water & Sanitation	680,319,943	15.4
3	Transport	493,091,138	11.2
4	Education	447,543,037	10.2
5	Other - Multisector	310,906,296	7.1
6	Social Infrastructure & Services	216,981,402	4.9
7	Industrial Development	127,000,000	2.9
8	Government & Civil Society	112,059,837	2.5
9	Agriculture	100,261,378	2.3
10	Land & Property Rights	88,087,579	2.0
11	Basic Health	52,989,063	1.2
12	Business Development Services	42,828,240	1.0
13	Environmental Protection	31,415,478	0.7
	Total	4,405,686,864	100.0

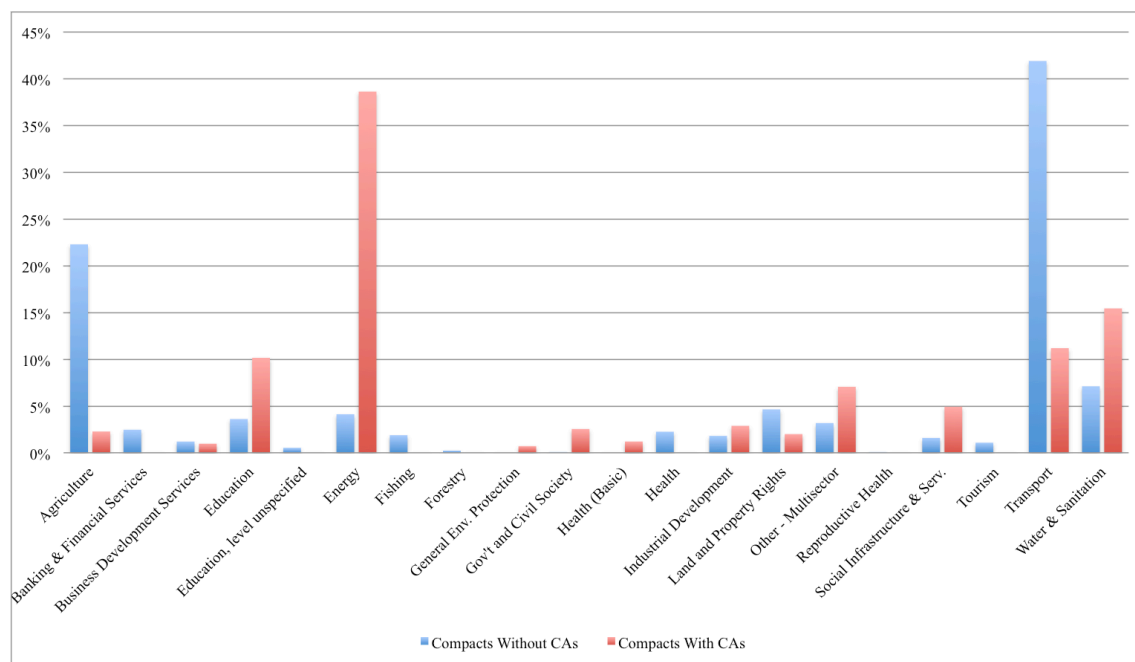
*Source:* Compacts' Summaries of Multi-Year Financial Plans.

*Notes:* All dollar figures in 2015-dollar terms.

**Figure 5.10** shows the difference between compacts designed with and without CAs. The top four sectors that made up 75 percent of total investments for the “Without CA” cohort were Transport (42%), Agriculture (22%), Water and Sanitation (7%) and Land and Property Rights (5%). For compacts with CAs the top sectors were Energy (39%), Water and Sanitation (15%), Transport (11%), and Education (10%). Transport continued to be among the top four, but MCC invested 74 percent less money in transport in compacts designed with a CA. Meanwhile investment in the Water and Sanitation sector, also among the top four, doubled. The biggest changes were in the ARD and energy sectors, which swapped places. After the CA, MCC essentially stopped investing in agriculture, dropping its investment in the sector by 90

percent. Meanwhile, energy increased by 90 percent. Sectors such as Fishing, Forestry, Tourism, and Reproductive Health disappeared altogether.

**Figure 5.10.** Distribution of Funding Across Sectors Between Compact Cohorts



*Source:* Compacts’ Summaries of Multi-Year Financial Plans.

**The Power Africa Program and New Forms of Bias.** Did this shift away from the agriculture and transport sectors mean that these sectors ceased to be constraints to growth? Or, to put it another way, were country governments able to see the energy and water sectors as major constraints to growth only because of the CA? The answer to both questions is “no”. At the turn of the decade, agriculture and rural development became a priority among countries in the Global South, particularly among governments in Sub-Saharan Africa. A senior MCC agricultural specialist said, “There was a demand from the governments of compact countries...they had a significant problem in the rural areas. They wanted to show something was being done in the rural areas and MCC resources gave the countries some very specific

projects to address rural poverty and rural migration to the cities” (Interviewee 2P, personal communication, March 29, 2016). This specialist was echoing the appeals of African leaders for the US to increase investments in agricultural production through the MCA (Chissano, Kufuor, & Mcpherson, 2002). This priority dovetailed with civil society organizations’ focus on the poor. The development finance institutions’ PRSP process allowed these groups’ joint concerns to translate into development and poverty reduction proposals and projects in a sector that had been neglected for two decades through the 1980s and 1990s (Bezemer & Headey, 2008).

As part of MCC compact development, the CA blunted this process. The CA was unable, because of the data it relied on and its practitioners’ orthodox economic proclivities, to perceive constraints in the rural economy. Of the 14 compacts developed with a CA, only one, the Moldova compact, which was the first to be developed using the CA methodology, had any substantial focus on agricultural and rural development. However, this was in spite of rather than because of its CA report; it did not identify irrigation or other rural development priorities as binding or nonbinding constraints. Constraints analyses for only three countries—Indonesia, Malawi, and Georgia—identified irrigation as a constraint (all nonbinding). None of these countries’ eventual compacts invested in agriculture and rural development. The analytical invisibility of binding constraints in the rural economy made it difficult to argue for compact investments in agriculture and rural development (Interviewee 2G, personal communication, February 26, 2016).

At the same time, CAs changed the nature of bias in decision making. They created space for senior officials’ bias at MCC headquarters to factor into investment decisions. Pressure from Washington always played a role in compact development. Proposals often changed following negotiation with MCC headquarters. However, in earlier compacts, partner countries initiated

and led the proposal process drawing on the PRSP and other planning documents and activities. With the CA, while the partner country's economics team was technically in charge, in practice they followed the lead of MCC economists (and their consultants) who understood the CA far better than they did. The CA shifted the locus of compact development activity and thus negotiating capital back towards MCC headquarters. This facilitated the entry of White House politics into compact development and partly eroded MCC's commitment to country ownership.

On average, CAs identified three binding constraints (Table 5.5). This was because, according to an MCC economist, "A single constraint would pigeon-hole the institution into doing something in that constraint without any room to maneuver" (Interviewee 2M, personal communication, March 22, 2016). The absence of a ranked prioritization among identified constraints left room, albeit more limited than before, for old ways of negotiating, or bias, to enter: "You could drive a truck through a lot of [constraints analyses]. And it points you in directions, but then there's a lot of discretion for decision making within those directions" (Interviewee 3C, personal communication, May 26, 2016).

The bias in Washington, DC at that time was towards the energy sector. In two of MCC's earlier experiences with CAs completed in 2011—Ghana and Tanzania—energy was identified as a binding constraint to growth. According to Eric Postel, USAID's Associate Administrator, the Ghana and Tanzania experiences led to the Obama Administration's Power Africa initiative (personal communication, July 28, 2015). Power Africa was designed to enhance power sector institutions, address constraints to investment, and enhance energy sector governance. According to a Congressional Research Service Report, twelve US agencies participating in Power Africa have committed up to \$9.7 billion to the initiative through 2018 in

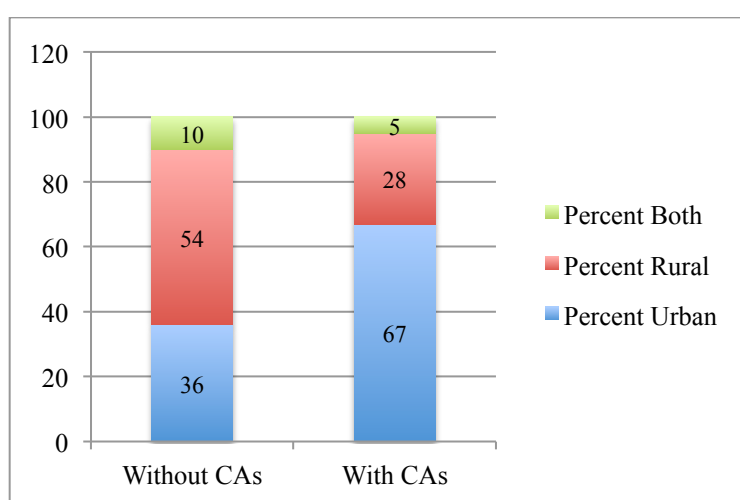
the form of technical aid, grants, export and trade capital and risk mitigation tools, loans, and other resources (Cook, Campbell, Brown, & Ratner, 2015).

Power Africa affected MCC's future compact decisions. Rather than the "strong ministry" culture in partner countries influencing which sector to invest in, the White House exerted pressure on agencies involved in Power Africa. In a situation where multiple constraints were identified and one was not prioritized over another, the bias was towards energy. One sector specialist remarked that "[t]he pressure to align MCC with Power Africa led senior officials to ignore other sectors" (Interviewee 2L, personal communication, March 22, 2016). For example, in Liberia both energy and transport had been identified as binding constraints to growth and the country government argued for a compact focused on road development. Nonetheless, the energy sector dominated the compact because of Power Africa (Interviewee 2T, personal communication, April 5, 2016). Of the 14 compacts developed with a CA, seven are in Sub-Saharan Africa. All but one, Cape Verde's second compact, focused on the energy sector. Meanwhile, of the seven compacts not in Sub-Saharan Africa, only Indonesia's compact had a focus on energy and it is relatively minor.

**The Spatial Distribution of Development Funds.** In addition to affecting which sectors MCC invested in, CAs affected the spatial distribution of development finance. Two-thirds (67 percent) of all committed funds in compacts developed with CAs were allocated to activities located in or benefiting urban areas (see **Figure 5.11**). Before CAs, compact funds were invested in urban areas 36 percent of the time. This represents a 60 percent difference. Meanwhile, the MCC invested 54 percent and 28 percent of funds in rural areas in compacts developed without and with CAs, respectively. This decline is not just because of the shift away from the agriculture sector. Prior to MCC's adoption of CAs, MCC had large programs in,

among others, rural water (Mozambique, Lesotho, and Ghana I), rural property rights (Burkina Faso, Madagascar, and Nicaragua), and rural roads (Cape Verde, Tanzania I, and Armenia). Constraints analyses not only shifted funds across sectors from Agriculture and Rural Development and Transport to Energy and Water and Sanitation, but also within sectors from investments in rural areas to urban areas.

**Figure 5.11.** Percent of Total Compact Funds Allocated to Urban or Rural Activities



*Source:* Country Compact Descriptions and Summaries of Multi-Year Financial Plans.

*Notes:* MCC does not designate a spatial classification to its activities. The classifications “urban”, “rural”, and “both” are subjective assessments applied to each activity MCC has funded between 2005 and 2015.

The reason we see greater urban programming after CAs were adopted is because of a combination of the methodological features of the CA, such as its use of shadow prices, and economists’ greater influence over project design and selection. The constraint analysis correlates GDP and other measures of capital formation with service levels and implies that capital formation as a result of infrastructure levels is a desirable outcome. Thus, infrastructure must be delivered in a way to maximize capital formation. Rural areas exhibit relatively lower “impulse responses” or changes in wealth, investment, and agglomeration to infrastructure investments. In order to boost aggregate investment, infrastructure will be directed to the urban

center rather than rural peripheries, particularly if cost-recovery of service delivery is a governing principle. A senior MCC official put it this way:

In [Country] there is a focus on moving goods and people as a constraint to growth, but the constraints analysis considers this constraint to be within the city confines. Some have argued that the transportation constraint should be considered more broadly, say across the country. But the constraints analysis and economists focus on the capital because it is the economic hub; it has the main port (Interviewee 2H, personal communication, March 2, 2016).

Some economic geographers claim that supposedly space-neutral approaches such as growth diagnostics are implicitly “capital-city promoting” (McCann & Rodríguez-Pose, 2011). The constraints analysis methodology combined with MCC economists’ Finance Ministry tendency favors the center over the periphery. This in turn disregards agricultural and rural development interventions. According to a social and gender assessment specialist who was skeptical of the CA’s narrow focus on private investment and formal firms, “[Growth diagnostics] is about growth in the center” (Interviewee 2G, personal communication, February 26, 2016).

**Shifts Towards Policy and Institutional Reform.** Finally, in addition to CAs changing the distribution across and within sectors, CAs changed the nature of projects as well. They shifted funding away from hard investments in infrastructure towards investments in policy and institutional reform. MCC has been interested in reform as part of its investments since the beginning. However, MCC did not have the appropriate instruments for carrying out this type of work.<sup>37</sup> After the CA, the typical compact was focused on a single sector with only three

---

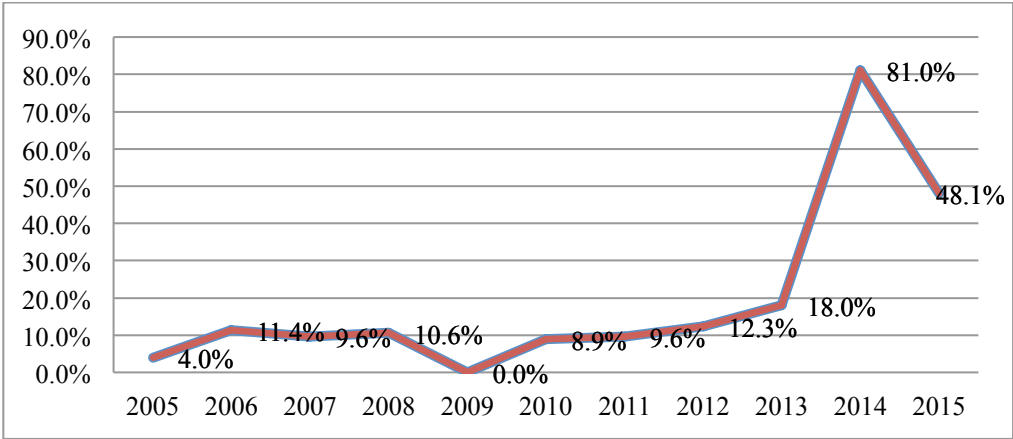
<sup>37</sup> Before CAs MCC relied on something called “Conditions Precedents”. Conditions precedents were, as the term implies, conditions that countries would need to fulfill before receiving a disbursement of funds, which occurred every fiscal quarter. Conditions precedents would be agreed upon when compacts were signed and written into the disbursement schedule based on a project timeline. Many conditions precedents involved policy or institutional changes like the establishment of a road maintenance fund. The partner country government would have to take action on something like this before funding for a highway rehabilitation project would be disbursed. The issue with conditions precedents was that they were often part of checklists that lacked “teeth”, were underfunded, and not always strategically integrated into compact design.



projects and at least one of those projects was usually focused on reform. Investments in reform involve funding for changes to policy, laws, regulations, and the institutions responsible for implementing them.

Constraints analyses helped change MCC’s work on reform. MCC now does reform “much more, with much greater ambition, and much greater confidence” (Interviewee 2P, personal communication, March 29, 2016). According to one senior official, “That’s not only because of the constraints analysis, but it was an important part of it” (Interviewee 2P, personal communication, March 29, 2016). Constraints analyses were central to this shift because they “allowed a modeling of policy change as a constraint to growth. The constraints analysis helped model policy as a first order constraint and allowed for a comparison of its benefits alongside typical ‘hard’ projects like roads” (Interviewee 2B, personal communication, February 11, 2016). Compacts developed without a CA invested 9 percent of their funds to reform related activities. This jumped to 33 percent among compacts developed with a CA, which represents an almost four-fold increase (see **Figure 5.12**).

**Figure 5.12.** Percent of Funds Going to Reform Activities Before and After CAs Affected Compact Design (2010)



Source: Compacts’ Summaries of Multi-Year Financial Plans.

As a result of the increased focus on reform, MCC's compacts increasingly resembled World Bank Sector Adjustment Loans, or SECALs. Sector adjustment loans were part of the World Bank's structural adjustment lending in the 1980s and 1990s. In 1982, the SECAL was introduced for policy changes and institutional reforms in a specific sector (World Bank, 2001). They focused on major sectoral issues such as the incentive and regulatory frameworks for private sector development, institutional capability, and expenditure programs. Between 1980 and 2000 the World Bank operationalized 233 SECALs worth \$42.5 billion, making up 44 percent of total structural adjustment lending (World Bank, 2001). This is how a World Bank energy SECAL in Senegal approved in 1998 for \$100 million was described: "This program promotes reforms in the energy sector, with the aim of providing more efficient electric power service and lowering electricity and petroleum prices. These reforms—which support the government's medium-term reform agenda—will reduce factor costs, improve Senegal's competitiveness and growth prospects, and increase job opportunities" (World Bank, 2001: 15).

The challenge is that MCC is not a development bank, but a bilateral grant-making agency. MCC's embrace of growth diagnostics—a tool that sought to displace a laundry list approach to economic *reforms*—to narrow down laundry lists of proposed *projects* altered the agency's portfolio to look more like the World Bank's structural adjustment lending in the 1980s and 1990s. While many at MCC wanted a tool to narrow down proposed projects, it's not clear they intended to change the fundamental nature of MCC's programs. A senior MCC official who had worked for years at the World Bank, including on SECALs, put the change in compact composition in a broader context:

We were set up in order to do what's called a developmental intervention in an eligible county for five years and that intervention alone would contribute to economic growth and poverty reduction. That was the whole purpose of creating MCC. The purpose of MCC is not to solve a constraint that is inhibiting growth at a macro level... So, why are

we focused on that when the whole purpose of creating MCC was project based? (Interviewee 2T, personal communication, April 5, 2016).

## **Conclusion**

As the fields of development policy and development economics reckoned with the ramifications of the Washington Consensus, three Harvard University economists developed a methodology to diagnose what prevented greater private capital investment in countries. The approach—rooted in contextualized problem-solving—was a direct response to the universal prescriptions of the Washington Consensus period. MCC economists proposed adapting and adopting this method to guide investment decision-making for MCC compact development. Their concern was that existing approaches were too politicized and therefore inefficient, both in terms of organizational processes and economic growth outcomes for countries. They successfully enrolled senior officials in their efforts at enhanced engagement with country core-teams. Growth diagnostics were introduced at MCC with little to no resistance despite being at odds with the way many MCC environmental and social assessment experts conceived of development. This was principally because these specialists did not, at the time, present a solution to senior management of their own, express their opposition as a coherent rationale, or organize themselves into an opposing network.

Growth diagnostics reshaped the compact development process by putting economists at the forefront of compact development, including the consultative process. Expressed through various features of the diagnostic such as the use of global competitiveness surveys and tests employing shadow prices, their economic rationality influenced the composition of compacts. A comparison of compacts developed with and without growth diagnostics illustrates the method's role in shifting funds away from the agriculture and rural development sector and into the energy sector; rural areas and into urban areas; and hard infrastructure into policy and institutional

reform investments. In half a decade, the deployment of a tool designed to reform structural adjustment lending practices in an aid agency with a project finance model made MCC's investments look more like the structural adjustment programs one might find at multilateral development banks in the 1990s.

## Chapter 6: The Power of Randomized Controlled Trials in Development Projects

### **Introduction**

This third case examines MCC's use of randomized controlled trials (RCT) to conduct impact evaluations of development projects. MIT's Poverty Action Lab made RCTs popular by the early 2000s. Around that same time, the Center for Global Development began to review how the development community conducted project evaluations. This chapter describes how a network formed around RCTs consisting of the Poverty Action Lab, the Center for Global Development, and the Bush Administration's interagency group working at the time to create the Millennium Challenge Account. It then explains how MCC evaluation economists, with help from this network, transformed the academic purposes of the RCT for development evaluation and got the Millennium Challenge Corporation's (MCC) Monitoring and Evaluation department to adopt RCTs despite resistance from more "traditional" performance evaluators and sector specialists. By 2015, MCC had initiated 23 RCTs, accounting for one out of every six impact evaluations at the agency. RCTs were not, however, distributed evenly among MCC's investments and this chapter lays out the reasons for this imbalance.

This chapter then discusses how MCC's adoption of RCTs affected project participant selection and implementation on a subset of farmer training projects. It focuses on how randomization and the need to limit contamination between treatment and control groups conflicted with project implementers' existing rationales for project participant selection and execution. It presents evidence of how project participants and implementation were different as a consequence of RCTs. The chapter closes with a discussion of what the outcomes were from some of MCC's earliest RCTs in farmer training projects and what those outcomes meant for farmer training projects and MCC's results agenda.

## **The Rise of RCTs in Development Economics**

Perhaps no other figure is synonymous with the explosion of RCTs in development as Esther Duflo, MIT Professor of Poverty Alleviation and Development Economics. Duflo, originally from France, graduated with a doctorate in economics from the Massachusetts Institute of Technology (MIT) in 1999. The MIT economics department is considered one of the most prestigious and influential economics departments globally. It has produced at least 14 Nobel Laureates and has a storied history. Paul Samuelson, who fused economics with mathematics, was a key figure in establishing the department and Robert Solow, the “father” of growth accounting and modeling, began his career there in 1949 and stayed until reaching emeritus status in 1995.

Esther Duflo’s dissertation *Essays in Empirical Development Economics* redefined an academic field in search of a new identity. Supervised by Abhijit Banerjee, who co-authored one of her dissertation chapters, Duflo studied three cases of development: primary school construction in Indonesia, social pensions in South Africa, and reputation’s role on contractual outcomes in the Indian software industry. Given how many dissertations in development economics today are based on RCTs, it is perhaps surprising that the method’s chief proponent did not employ any in her own. Nonetheless, Duflo’s Indonesia chapter spearheaded the RCT movement. This is because she relied on the Indonesian government’s decision to build 61,000 primary schools across the country between 1973 and 1978 as an instrumental variable. She called this program an “unusual policy experiment” because it both drastically changed the Indonesian government’s expenditure levels on primary education and affected a specific population of students: children younger than 13 in 1973. By using the variation in school expenditures generated by this policy as an instrumental variable for the impact of education on

wages, Duflo was able to isolate and estimate the program's economic returns to education using Indonesian census data (Duflo, 1999).

The dominant method in development economics research at the time was cross-country growth regressions. In the 1980s and into the 1990s, development economics researchers and doctoral students preferred conducting growth regressions to identify the successful determinants of economic growth (Durlauf, 2009). Through deduction and stylized models, they attempted to explain the determinants of growth empirically (Milberg, 2009). These regression studies, however, invited critique for their failure to establish causality. Analysts were accused of explanatory variable selection bias for the production and publication of statistically significant results (Eble, Boone, & Elbourne, 2017).

Duflo's dissertation addressed the identification problem for outcome variables using a natural policy experiment, which was highly influential. This overshadowed the fact that she had done so only for a specific sectoral program rather than national level economic policy, which was the usual focus of growth regressions. The important thing was that she had established causality in a way that bypassed some of the challenges of identifying an instrument in growth regressions. As a technique for isolating causality, randomization was considered far superior and preferable to instrumental variables. But it was not until she and her former advisor, Abhijit Banerjee, began working with nongovernmental organizations (NGO) in India in the early 2000s that the idea to use random allocation of program interventions to test impact really began.

Duflo and Banerjee partnered with Pratham and Seva Mandir, two NGOs working on education in India. Pratham was much larger and worked in several states while Seva Mandir was small and had operations only in Rajasthan state. The key lesson they took away from these experiences was that partnering with development organizations allowed researchers to modify

and test micro-interventions, such as extra tutors or blackboards in classrooms, through pilot programs to discover “what works”. The quality of the partnership and its limited scale allowed randomization and the implementation of RCTs.<sup>38</sup> Experimental results could then be built inductively without, supposedly, the imposition of theory.

This argument for rigorous experimental results was appealing to a field that was in search of legitimacy within the academy (Bhatt, 2011; Porter, 1996). The fact that external development organizations were willing to partner with development economists and fund research was all the more helpful. A decade and a half later, RCTs became a mainstay within American economics departments. Mainstream development economics has been gradually decimated into “micro-level impact appraisal studies” of developmental projects and the broader macroeconomic questions have been displaced from the field (Akbulut et al., 2015).

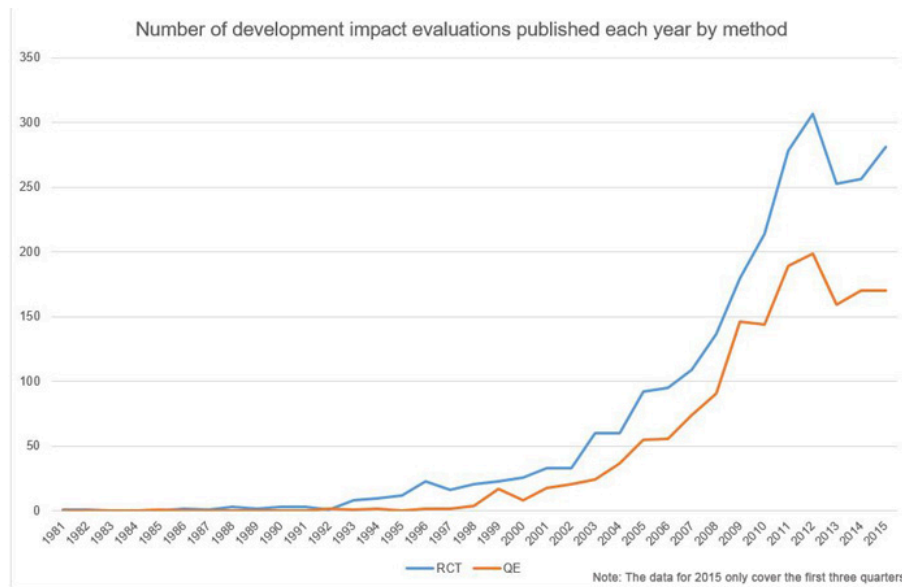
In 2003, Duflo and Banerjee institutionalized their approach to development economics by co-founding the MIT Abdul Latif Jameel Poverty Action Lab (J-PAL). J-PAL’s founders say its mission is “[t]o reduce poverty by ensuring that policy is based on scientific evidence” (Banerjee & Duflo, 2012). By early 2016, J-PAL had initiated 732 RCTs on five continents. It has already completed 485 of those (J-PAL, 2016). J-PAL has six global offices and spurred partners committed to its mission, namely Innovations for Poverty Action at Yale University and the Center of Evaluation for Global Action at the University of California Berkeley. It would not be an overstatement to suggest that J-PAL and its many academic affiliates have transformed what counts as development knowledge. There has been an explosion of published studies using the RCT methodology (see **Figure 6.1**):

---

<sup>38</sup> Throughout this chapter I will use the term RCT to refer to a particular form of randomized trial conducted in non-laboratory social settings, which some scholars prefer to call a randomized field experiment (RFE).



**Figure 6.1.** The Rise of RCTs in Development



Source: 3ie (2015). 3ie (2015). *Impact Evaluation Repository*. Available: <http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/>

Notes: QE stands for quasi-experimental designs (i.e., a counterfactual is established without randomization).

These research centers also regularly provide training and consulting services to major development, philanthropic, and academic institutions across the globe. The surge in RCTs has had a major impact on the practices of multilateral development institutions such as the World Bank, philanthropic foundations such as the Bill and Melinda Gates Foundation, bilateral donors such as the German Corporation for International Cooperation, and international nongovernmental organizations. By 2010 the World Bank alone had initiated over 200 RCTs (Legovini, 2010). As one review of the field put it, “development economics can best be understood as...responding to the research, project evaluation, auditing requirements of the national and international, governmental and nongovernmental institutions of development” (Akbulut et al., 2015: 749).

## **Stabilizing RCTs in MCC**

While the empirical turn in development economics throughout North America paved the way for the rise of RCTs, how did their popularity in academia translate into their widespread adoption among various development agencies? How did changes in development economics dovetail with changes in development policy and administration to make RCTs an attractive governmental rationality? In 2004, the Center for Global Development (CGD) set up an evaluation “working group” to review how the development community conducted project evaluations. The working group established standards of evidence. Based on these standards, it determined that there was a significant “gap” between using counterfactuals to assess impact—of which RCTs represented the “gold standard” approach—and evaluations that development agencies were conducting. A network formed around RCTs consisting of J-PAL, the CGD evaluation working group, and the George W. Bush administration’s interagency group that created the MCA. With the help of CGD’s establishment of evaluation evidence standards, MCC evaluation economists transformed the academic purposes of the RCT for development evaluation.

**Mind the Gap: RCTs and Accountability Converge Around Impact Evaluation.** In the late 1990s a select group of professional economists, or economists “in the wild” (Callon & Rabearisoa, 2003), and development elites began to grow concerned about identifying developmental impact. Their principal complaint was that the development community was not learning, specifically, about whether their investments “worked”:

The frustration I’m referring to is one that I have personally felt over the years working in the development business, being asked to design, implement and advise developing country governments on complex development programs in the health and education sectors with little more to go on than some theory, my own on-the-job experience and observations, some good advice from colleagues and the occasional case report in the literature...\$30 billion of development assistance are spent annually...on well-

intentioned development programs that are based on very weak evidence about what works (When Will We Ever Learn, 2006: 2).

The quote above is from Ruth Levine, co-author of a report of the Center for Global Development called *When Will We Ever Learn: Improving Lives Through Impact Evaluation* (“The Report”) (The Evaluation Gap Working Group, 2006). Levine and fellow colleague William Savedoff convened the CGD Evaluation Gap Working Group. The 21-member Working Group was comprised of 10 professional and academic economists, five development policy and four health sector elites, and two evaluation professionals. Esther Duflo, co-founder of MIT’s J-PAL, was among the economist members. The Working Group was financed by a grant to CGD from the Bill and Melinda Gates Foundation and the William and Flora Hewlett Foundation to, according to the Report, “Address the lack of information about the effectiveness of social programs in low- and middle-income countries” (The Evaluation Gap Working Group, 2006: 46). The Working Group sought to understand the reasons for the dearth of good impact evaluation and the possible ways to make significant progress toward closing the “evaluation gap” (The Evaluation Gap Working Group, 2006).

The Working Group reviewed the current status of impact evaluations in social sector programs and developed recommendations to address the evaluation gap problem. In doing this it consulted with over 100 stakeholders, including governments, the research community, private foundations, multilateral and bilateral agencies, and major international NGOs. Among the 36 individuals closely consulted for the Report were Abhijit Banerjee, the other J-PAL founder, and Rachel Glennerster, J-PAL’s executive director (Center for Global Development, n.d.).

The evaluation gap in the Report refers to the space between where the state of development evidence is and where it ought to be. But what is good evidence? How does one define impact? And how does one go about determining if a program works? These are all

questions of epistemology, or how humans create knowledge. By answering them, the Working Group was establishing standards of evidence and creating a hierarchy of knowledge. The RCT as a gold standard in development impact evaluation came to represent one such system of authority. Thus, to identify and direct others' gaze towards a gap in knowledge requires, first, the construction of a hierarchy of knowledge, of elevating the value of some ways of knowing above others and, then, to suggest that the knowledge most worth knowing is currently in short supply. In other words, these economists cultivated a sense of ignorance and scarcity.

Prior to the Working Group's formation in 2004, there was no dearth of knowledge *per se* in the field of development. Even one of the report's authors, Ruth Levine, acknowledged this:

I think you know there's a kind of perpetual problem that we've encountered in this project with the word "evaluation", because it's really, it's such a broad word. It covers so many different kinds of knowledge generation. And what we tried in a sort of very disciplined way was to consistently talk about impact evaluation. But there's a whole range of other kinds of evaluations that are done within development institutions (When Will We Ever Learn, 2006: 20).

The Working Group identified multiple forms of "other kinds" of knowledge generated by development agencies: i) monitoring data, which verifies whether programs were implemented according to plan, expected outputs were realized, and intended beneficiaries were reached; ii) process evaluations, which ask how and why programs get implemented and generate "best practices" and "lessons learned"; iii) accounting data, which are usually compiled in financial statements as a way to account for taxpayer spending; and iv) impact evaluations, which generate knowledge about whether a program achieved its basic aims (The Evaluation Gap Working Group, 2006). But what is it about these other reports that make them necessary but insufficient? If internal evaluation teams contribute important knowledge to "good decisions", then why are we still not learning? Because, according to the Working Group, learning is synonymous with impact, which is invaluable to attribution of the outcome to the intervention.

A true impact evaluation must have a counterfactual. An evaluation that simply assesses the effect on beneficiaries before and after the program without reference to a comparison group is at best invalid if not entirely harmful and misleading: “To determine what works, it is necessary to collect data to estimate what would have happened without the program...By examining how people in a particular program fared *vis-à-vis* an appropriate comparison group, it is possible to measure the *impact* that can be *attributed* to the specific program. And it is only this impact that truly measures the program’s effect” (Svedoff & Levine, 2006: 2). Meanwhile, process evaluations “[p]rovide little insight into our actual impacts and, although crucial, their contribution to knowledge essentially focuses on a better understanding of operational constraints and local institutional and social contexts” (Jacquet, 2006: 178).

A valid counterfactual, or a state of the world without the program, is a construct, an epistemological product that “mimics” a world that in reality cannot be observed. Experimental and quasi-experimental methods are the best methods to achieve credible counterfactuals. The gold standard in experimental methods is the RCT. It doesn’t produce just any counterfactual, but a control group, whose outcomes when compared to those of the treatment group, or beneficiaries receiving the intervention, produces the most internally valid estimates of program impact. Other quasi-experimental methods, chiefly statistical matching, interrupted time series, and instrumental variable approaches all try to approximate the experimental assignment of treatment and control groups but all run into problems of biases related to observable differences between the two groups. The technical features of these methods are less relevant here than the recognition that there is an explicitly accepted hierarchy among them with the RCT comfortably resting atop.

The RCT’s superiority in constructing control groups favorable to attributing project

specific effects explains why the method, so popular with American economists in the academy, dovetailed with ambitions of economists in the wild to make impact evaluation practices among development agencies and policymakers more rigorous and scientific. But this still does not explain how they made their way, despite opposition, into development agencies' evaluation units and, in the case of the MCC, became the preferred evaluation approach.

**Belling the Cat: MCC as an Evaluation Pioneer.** In describing the challenge of getting development agencies to adopt impact evaluations, CGD's William Savedoff asked: "Who will bell the cat?" (When Will We Ever Learn, 2006: 10). *Belling the Cat* is a fable that provides the lesson of how easy it is to propose impossible remedies. A young mouse suggested to his fellow mice that belling a marauding cat would warn them of her approach. An old mouse replied, "That is all very well, but who is to bell the Cat?" (Jacobs, 1909). The dangers in development policy are not as grave but impact evaluations do present a collective action problem because of myriad disincentives: negative findings may lead to budget cuts or loss of organizational prestige; rapid start-up and implementation are rewarded over the time to conduct evaluations; and politicians value social programs to gain electoral advantage rather than long-term social goals (The Evaluation Gap Working Group, 2006). Thus Savedoff pleaded that "[w]e have to get pioneers who are willing to take the risk" (When Will We Ever Learn, 2006: 10). Among U.S. development agencies, MCC belled the cat.

MCC felt comfortable taking this risk in large part because of CGD's efforts to establish RCTs as the gold standard for project evaluation. The CGD Working Group was not just engaged in an intellectual exercise but in building a network through which impact evaluation gains power. As part of its consultations, CGD conducted two "roundtables" where the Working Group's objectives and ideas were discussed and roundtable participants were asked for

feedback. One of these roundtables was held at the William and Flora Hewlett Foundation in Menlo Park, CA on July 25, 2005 and the other at CGD in Washington, DC on August 1, 2005. Attendees of the CGD roundtable included Delia Welsh, MCC's Associate Director for Monitoring and Evaluation (M&E), and Andrew Warner, MCC's Senior Economist. Jon Baron, Executive Director of the Coalition for Evidence-Based Policy (CEP), an advocacy group that lobbies the federal government (including MCC) to use RCTs, also attended.

While the Working Group was busy setting standards and building an epistemic community (Haas, 1992) for impact evaluation, another working group was busy operationalizing the MCC to manage the Bush Administration's Millennium Challenge Account (MCA). The White House put together an interagency working group to establish the MCC while Congress reviewed MCA legislation. It included members from the Treasury and State Departments, Office of Management and Budget (OMB), United States Agency for International Development (USAID), and National Security Council (NSC) (see Chapter 4). In addition to "selectivity" (of countries it would fund) and "country ownership" (of investment plans), one of MCC's guiding principles was a "focus on results". Members from Treasury and OMB were responsible for defining results and to them it meant a reliance on RCTs to evaluate projects: "That's why OMB got so involved with MCC because it was a completely performance driven agency...and the RCT thing was really just because the space was wide open. In the typical Bush administration initiative they just had principles but no operational plans" (Interviewee 3K, personal communication, July 7, 2016). MCC working group members from Treasury and OMB went on to hold senior positions in monitoring and evaluation (M&E) at MCC.

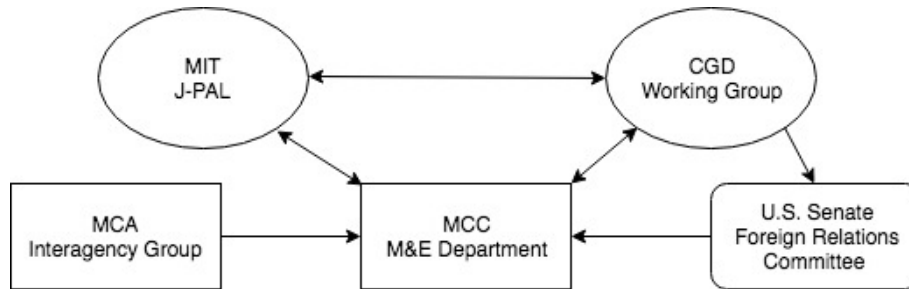
In the same year that the Working Group was conducting its consultations, J-PAL was preparing to launch its first ever Executive Training Course called *Evaluating Social Programs*.

Up until that moment, J-PAL had spent its energy training undergraduate and graduate economics students in the arts of field experiments, power calculations, and intent-to-treat estimates. Duflo, Banerjee, and colleagues were now ready to repackage the RCT for popular policy consumption. The executive training course ran for five days from March 21<sup>st</sup> to 25<sup>th</sup>, 2005 in Cambridge, MA and consisted of a combination of lectures, group work, and group presentations. There were seven lecturers: three from MIT, three from Harvard, and one from Princeton. Lecture topics included *Why Randomize? How to Randomize? Managing Threats and Policy Impact and External Validity* (Poverty Action Lab, 2005). Group work centered on analyzing early field experiments that J-PAL had conducted in India and Kenya and, more importantly, programs that participants were grappling with from their respective agencies. There were 24 attendees—ten from multilateral development banks, seven from academia or research institutes, three from developing country ministries, three from the UN and NGOs, and two from the US federal government. One of the US federal government attendees was Delia Welsh, Associate Director for M&E at MCC (Poverty Action Lab, 2005).

The CGD Working Group, MCA interagency group, and J-PAL were part of a network whose ideas traveled into the state via its members (see **Figure 6.2**). In 2008 MCC cohosted an event with the Council for Excellence in Government in Washington, DC called *Rigorous Evidence: A Key to Progress Against World Poverty?* At the end of that event, Jon Baron, the executive director of the Coalition for Evidence-Based Policy, acknowledged Welsh's contribution at MCC: "Someone who's in this room who really was a key force in MCC's focus on rigorous impact evaluations, helping to make it part of the whole enterprise. It's Delia Welsh... Could you just stand up for a moment?" (Rigorous Evidence, 2008: 51).



**Figure 6.2.** The RCT Network



*Source:* Author.

*Notes:* Arrows represent common working relationships and exchange of people and influence.

**Making RCTs an Organizational Reality.** Between 2004 and 2006 RCTs had become an evaluation policy reality at MCC. The RCT network’s members, connected to MCC through Delia Welsh, had the power to translate broad policy ideals such as focusing on results into administrative reality. One way expertise influences development is because its “mundane and material technologies make it possible to translate the ‘programmatic’ ideas of government into practice” (Justesen & Mouritsen, 2011). The next steps were to build material resources around RCTs.

One important feature to impact evaluation advocates was independence. The PROGRESA evaluation in Mexico was lauded for involving experts outside the Mexican government.<sup>39</sup> William Savedoff said, “External actors help with expertise with quality and with integrity...It was very important, for example, in Progressa, that the academics were involved, an external agent” (When Will We Ever Learn, 2006: 7). MCC adopted the PROGRESA model and the Working Group’s recommendation when it launched an ID/IQ (indefinite delivery / indefinite quantity) contracting vehicle to procure evaluation services from a select number of

---

<sup>39</sup> PROGRESA was a major conditional cash transfer and social welfare program that started in Mexico in 1997.

evaluation firms. When soliciting bids, the ID/IQ provided a Statement of Work that described what the firms would be required to do and requested firms' proof of prior experience. An MCC solicitation for proposals stated that: "[t]he contractor shall provide technical assistance in a variety of areas in carrying out evaluations using a randomized design in response to long-term and short-term task assignments. Experience with the methodology of randomized trials is required as well as application of the methodology in developing countries" (MCC, 2006c: 6).

The ID/IQ contract represented a unification of the Working Group's recommendations and MCC evaluation guidance. MCC was no longer just trying to bell the cat for symbolic purposes but was putting material resources behind RCTs. Human resources were also geared, in part, towards RCTs. In the words of one early M&E manager, "Every time we did recruiting, [experience with RCTs] is what we looked for" (Interviewee 3B, personal communication, April 27, 2016).

### **Conflicting Rationalities in Participant Selection**

To compete, a particular intellectual rationality such as RCTs must contend with other rationalities, and, if it is to survive and thrive, overcome them. There are three sites of conflict in the context of MCC that warrant examination: the Department of Accountability, Transaction Team, and process of compact implementation. Evaluation economists had to overcome peer dissent in the Department of Accountability from M&E staff with decades of experience in performance evaluation.<sup>40</sup> In the Transaction Team, they had to overcome sector opposition to RCTs. In compact implementation, they had to overcome opposition to RCTs from implementers in the field. Strategic and productive power was central to overcoming this

---

<sup>40</sup> Evaluation economists at MCC were often not PhD economists or trained exclusively in economics. They usually had masters' degrees in public policy, international development, or related fields and had training and skills in quantitative and data analysis and research design.

opposition.

**Overcoming Conflicting Rationalities.** RCTs have their advocates, but also their detractors. Arguments against RCTs are made most commonly for epistemological, operational, and ethical reasons. The most common methodological critique is that RCTs lack external validity. While particular experiments might generate internally (to the experiment) valid impact estimates, what is to say such estimates can be reliably transported to other environments (Deaton, 2010; Mookherjee, 2005)? This has been a longstanding criticism of RCTs that extends beyond its use in development economics and remains unsettled.

Many development practitioners are opposed to RCTs for operational and reputational reasons. There is intense pressure on project managers and implementers to disburse funds, deliver outputs to project participants, and do no harm. The MCC, for example, has been pressured to disburse funds quickly and admonished for not doing so (Dugger, 2007). Development agencies increasingly face scrutiny and backlash over social and environmental harm (Fox & Brown, 1998). The need to account for an RCT among the long list of other priorities is, in the short term, viewed as a nuisance to be avoided. In the long term, there may be disincentives to RCTs and impact evaluations more generally to minimize reputational risk. Few agencies and managers are comfortable being responsible for their projects' impacts when they have limited control over project implementation and virtually no control over other exogenous variables (e.g., weather patterns, input prices).

Another relevant challenge for development actors is the ethics of RCTs. It is ethically suspect, especially in resource poor environments, to deny a group of people access to a resource that potentially benefits them (Reddy, 2012; Ziliak & Teather-Posadas, 2014). This is where RCT advocates' exercise of establishing standards of evidence and knowledge is crucial. The

benefit of denying some participants resources is the knowledge, or learning, that can be gained from experimentation. Such knowledge is framed as a public good that will extend well beyond the specific project to benefit the entire development community (Stiglitz, 1999). But this definition of knowledge and evidence was created within a particular epistemic community that others will likely not share. If knowledge definitions are particular to epistemic communities, then other actors, such as project managers, sector specialists, and implementers, who are trained within other epistemic communities such as engineering, agronomy, or public health will have different standards of evidence and knowledge. More importantly, they will have other ways and means of knowing based on their accumulated development experience. They will not be convinced, at least at first, that they are ignorant of what works. These competing knowledge claims, between the evaluation economists' quest for truth through epistemic method and the development practitioner's accumulated knowledge through experience, sets up "conflicting rationalities".

**Sites of Social Control.** According to Breslau's study of RCTs in United States' labor market and welfare policy, the success of RCTs is rooted in social control:

An experiment is more than just a technical problem of research design but is also a negotiation of political terrain... The integrity of the experiments was preserved based on the degree of cooperation from operating staff, such as street level bureaucrats, and clients—Such cooperation requires a resolution of social control... When full social control is achieved, experimenters can even become the implementers (Breslau, 1998: 73).

The concept of "impact" was one of the crucial devices at evaluation economists' disposal to overcome opposition from peer dissenters and sector specialists. Impact evaluations, and thus RCTs, are essentially about accountability (after all, MCC's evaluation division was initially housed in the Department of Accountability). Ruth Levine made the point rather explicitly before the Senate Committee on Foreign Relations on the issue of *Promoting*

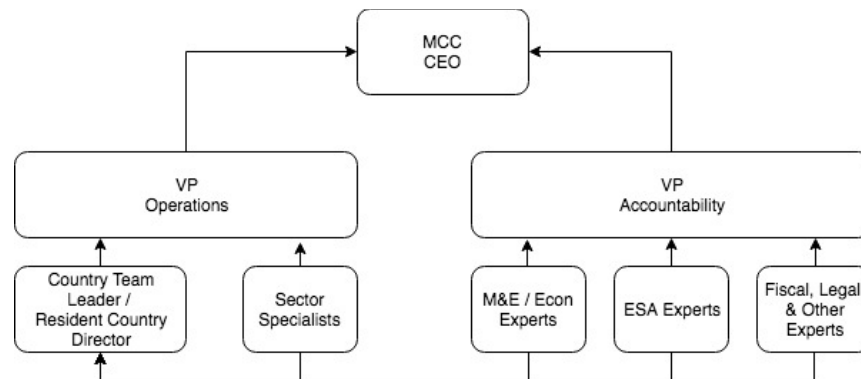
*Effectiveness and Fighting Corruption* among multilateral development banks: “Knowing whether or not the banks are succeeding in financing programs that directly improve people’s lives is the core of accountability” (Multilateral Development Banks, 2006: 1). And for Levine accountability was essentially about attribution to a discreet project or program: “Assessing this type of performance comes from impact evaluation, defined as evaluations that...can be directly credited to the program itself, as distinguished from changes that are due to other factors. That is, they are evaluations that permit attribution of program-specific effects” (Multilateral Development Banks, 2006: 2).

While impact evaluations may essentially be about accountability, the ability to cast them as either a tool for accountability *or* learning is why RCTs were smoothly stabilized within MCC’s evaluation division. This is because impact is a vehicular idea (McLennan, 2004; Osborne, 2004). Its purpose was to get the development community to “move on” from process evaluations that tell us a great deal about programs themselves to impact evaluations that will allow us to learn about program-specific effects. Evaluation economists could get behind impact because it meant *learning* what works in a way that made sense to them based on their training in economics or public policy. Senior officials that answered to MCC’s Board of Directors and Congress could embrace impact under the broad mantle and mandate for being *accountable* for results. Under the impact umbrella, senior officials and evaluation economists could coexist, benefit, and find meaning. As one senior M&E official put it: “People didn’t understand RCTs, but if it gives us truth with a capital ‘T’, that’s one of our three pillars; let’s go for it. [The Vice President of Accountability] didn’t understand RCTs, but he would support it because at that time results was a really big word” (Interviewee 3B, personal communication, April 27, 2016). With senior officials’ weight behind impact evaluations and RCTs, performance evaluators

within the Department of Accountability who had “ground level implementation experience” and were wary of the methodology, had little choice but to adjust to the new reality of RCTs.

Once RCTs and quasi-experimental methods were stabilized within MCC’s evaluation division and became official evaluation guidance, the big question facing evaluation economists was how to get reluctant sector colleagues on the Transaction Team responsible for program implementation to go along with an RCT. MCC was set up as a matrix organization (see **Figure 6.3**). This meant that teams were formed around different functions that had parity with each other. Each team had members from respective technical staff such as Environmental and Social Assessment (ESA), Sector Programs, Fiscal Accountability, Economic Analysis, M&E, Procurement, and other administrative and legal functions. The team was collectively responsible to a Country Team Leader (during compact implementation, a Resident Country Director was in charge) who reported to senior management. Each individual technical team member, however, was also accountable to his or her technical manager, who was a member of senior management.

**Figure 6.3.** MCC's Matrix Management Structure, 2004–2007



*Source:* Author, based on MCC organizational charts.

*Notes:* VP stands for Vice President. Vice Presidents were in charge of MCC Departments. Arrows represent reporting relationships.

In this arrangement the evaluation and sector specialists were equals. Having a “seat at the table” allowed MCC evaluation staff to achieve one of the Working Group’s key recommendations—co-designing projects and evaluations. Ruth Levine of CGD said, “Rigorous evaluation almost always requires integration into a project design... Even better, early integration of program evaluation and design often permits the rollout of a program to be randomized” (Levine, 2006: 25). MCC’s guidance reflects this commitment to “co-experimentation”: “Ideally, project design and evaluation design should be developed simultaneously” (MCC, 2006b: 20). Furthermore, co-experimentation was something J-PAL’s founders initially promoted. According to Abhijit Banerjee and Esther Duflo, “There was less emphasis on one-off evaluations, where the researcher is brought in to evaluate a specific program that the organization has already decided to evaluate... From the point of view of the researchers, this offered the possibility of moving from the role of the evaluator to the role of a co-experimenter, with an *important role in defining what gets evaluated*” (Banerjee & Duflo, 2008, emphasis added).

As mentioned earlier, there are substantial disincentives for program and sector specialists to engage in co-experimentation. For these reasons, sector specialists were reluctant to co-experiment and had to be convinced to do so; their interests had to be translated to match those of the evaluators. The main struggle between evaluation and sector specialists was whether and how to establish a control group. Thus, the main question became, “Is establishing a control group about learning or is it about accountability?” If it was about learning—within the definition of knowledge as a project-attributable impact along narrow measures such as income—then establishing a control group became more attractive. If on the other hand, it was about accountability—within the framework of knowing if impacts can be attributed to your

project rather than something exogenous like weather or sector specialists' own biases—then establishing a control group became a chore for sector specialists to avoid. Framing the identification of impact through establishment of a control group as learning was critical to gaining sector specialists' cooperation.

**Translating the Interests of Sector Specialists.** MCC's evaluation economists enrolled their sectors specialist colleagues in Washington and partner countries in the experimental cause by stressing the idea of impact evaluation as learning rather than accountability through MCC presentations and in-country workshops. A primary task under the ID/IQ contracting mechanism directs evaluation firms to conduct “working sessions or training with MCC staff and country counterparts...on the evaluation design” (MCC, 2006c: 7). It goes on to say, “Evaluations using randomized design will require that staff and country counterparts understand the methodology, are able to discuss why it is applied to specific projects, explain how the evaluations will be implemented, analyze results, and apply lessons” (MCC, 2006c: 7). MCC evaluation economists convened these workshops: they were between 1 to 3 days long, held in a country's capital, and led by a contracted evaluation expert, usually an economics professor. Evaluation economists' goal was to get people who would be involved with implementation to attend. A workshop in Ghana invited about 75 people—five from MCC, 17 from the Ghanaian core team, 20 from Ghanaian governmental ministries, 17 from other donor agencies, and 15 from the NGO and research communities (not all invitees attended) (MCC, 2006a). A workshop in El Salvador had 55 attendees across a similar spectrum of organizations (MCC, 2007a). The workshops' objectives were to “[d]iscuss evaluation topics such as selection bias, control groups, randomization, matching, and difference in differences. The purpose of the workshop is to train



[local counterparts] and implementing entity staff and to discuss possible impact evaluation designs for the El Salvador Program” (MCC, 2007b).

Evaluation economists believed the workshops were successful in achieving their objectives of teaching about impact evaluation, alleviating fears or concerns about randomization, engaging non-evaluation specialists in impact evaluation, and embedding evaluation design early on in program planning. One summary described the workshop experience in this way:

The impact evaluation workshop was successful and set the table for the upcoming meetings with each of the implementing entities. Although some parts of the presentation were somewhat technical for the audience, we received positive feedback overall that it was very helpful in clarifying many issues. The [attendees] were very interested in learning more about impact evaluation and became more receptive to the idea of randomization...In sum, [the economist] provided an expert, independent and unbiased view of possible designs for impact evaluations, which helped to answer many questions and also to alleviate concerns on the topic (MCC, 2007b: 1).

MCC is not the only organization that found evaluation workshops useful. In a blog post,

*Notes from the Field: the usefulness of early workshops*, World Bank economist Markus

Goldstein outlined why he likes doing an evaluation workshop, especially early in the process:

The next afternoon and the following morning are spent discussing the design of the impact evaluation. I try and keep this as broad as possible to bring the team in and see what kind of things we *might* do—this expansive list helps get people excited about what we are doing here...It’s also good to have a range of the folks implementing different components or parts of the program present...they’ll be critical for managing the integration of program implementation and the evaluation...A couple of other thoughts about why these workshops are useful. First, this really increases the program team’s ownership of the evaluation—they see why it’s useful and they are excited about the questions we’ll be able to answer. This is likely to help boost the chances that the evaluation won’t flop—if the team understands the logic of why you are (for example) randomizing in the first place, they’re probably less likely to undo the randomization (Goldstein, 2011).

The World Bank ran over 25 workshops in a five-year period. From fiscal years 2005 to 2010, over 3,100 people were trained on impact evaluation. Out of the 167 teams that participated in

the workshops, 144 implemented their impact evaluation for an “efficiency rate” of 86 percent (Legovini, 2010).

The ultimate goal of workshops was to transform potential adversaries into stakeholders in the learning enterprise. Workshops bonded participants together around a collective identity—experimenters. The economics professors from esteemed universities that led MCC workshops served as important sources and figures of authority in this effort. The focus on learning, rather than accountability, made the establishment of a control group much more palatable.<sup>41</sup>

The outcome of these MCC workshops, conducted in virtually every MCC Compact country, was often a tentative agreement on which program activities would be evaluated using experimental, quasi-experimental, or non-experimental methods. This agreement was later finalized through follow-up meetings and often resulted in legally binding language on the integration of impact evaluations into compact implementation. Specific design features, such as randomization, could then be written into contracts between the MCA and implementers. Through its experience with impact evaluations, MCC would later conclude that “aligning incentives” was central to successfully implementing RCTs (Farley, Lucas, Molyneaux, Penn, & Hogue, 2012). Part of aligning incentives was “creating clarity and incentives” through contracts: “Final contracts should, at the very least, ensure that implementers do not have incentives to undermine the evaluation methodology, and ideally creates incentives to actively support it” (MCC and USAID, 2011: 7). Nevertheless, as will be shown later, aligning

---

<sup>41</sup> Learning was also cultivated at MCC headquarters through less routine MCC presentations and guest speakers, usually from J-PAL.

incentives through contractual authority was a necessary but insufficient form of power for a successful RCT.

### **RCTs and Biases in Project Evaluation**

**RCTs at MCC.** Successful social control aligned the priorities of evaluators and sector specialists but did not determine which projects were chosen for an impact evaluation or RCT—this was largely a feature of the RCT itself. According to RCT proponents, the first and foremost reason to do an RCT over and above a quasi-experimental design is selection bias.

Randomization ensures that there is, on average, parity between treatment and control groups.

This is how one MCC evaluation economist pitched RCTs for a farmer-training program to her MCC sector colleagues based on an a J-PAL explanation of how randomization can determine if schooling improves health:

For the [country] impact evaluation, we want to randomize, mainly because we worry about selection bias. For example, how do we know if one group of extension service beneficiaries is applying new technology, growing [fruit], and increasing their incomes as a result of the extension service or because this special group is particularly entrepreneurial and driven? Are we biasing our results because we are selecting beneficiaries with high potential? We may attribute their success to the extension service, without really knowing if it was the service or something unobservable about these beneficiaries. Using selection criteria that is unrelated to potential success (i.e., randomization) can help us address this problem (Beatty, 2006).

A major selling point for random allocation is that such biases are dealt with through the equal chance of being allocated to the treatment or control group. Evaluation economists believe this ensures that any difference in outcomes observed after the project between the groups is a function of the intervention and not other “unobservables”.

While proponents regularly tout this benefit, they rarely ever speak of the structurally obscured biases of RCTs themselves. Several economists have pointed out how RCTs distort the research agenda in development economics because the method is not appropriate for all

questions (Barrett & Carter, 2010; Basu, 2005). There is, however, little theorization or evidence of this distortion or bias in development economics or policy. It is important to know why the RCT method is directed towards what Barrett and Carter (2010) call a “non-random subset of relevant topics” and what the implications of that are.

Over the decade spanning 2005–2015 MCC initiated over 150 evaluations. Of these, the majority were non-experimental or performance evaluations. About two out of every five evaluations were impact evaluations. When it came to coverage of MCC’s total investments portfolio in dollar terms, experimental evaluations, which include quasi-experimental designs, were on par with performance evaluations (see Table 6.1). Only 15 percent of MCC’s portfolio was not under evaluation at all or had an evaluation whose design was undetermined.

Table 6.1

*MCC's Evaluation Record as of 2015*

Activities with:	FY 2014–2015	
	N	Percent of total MCC investments covered (%)
Non-experimental evaluations	87	46
All experimental evaluations ( <i>quasi</i> + <i>RCT</i> )	57	42
<i>RCTs</i>	23	NA
Sub-total (non-experimental + experimental)	144	
<i>Experimental evaluations among subtotal (%)</i>	40	
<i>RCTs among experimental evaluations (%)</i>	40	
<i>RCTs among subtotal (%)</i>	16	
Evaluations with undetermined design	8	8
Total Evaluations	152	NA
Activities not under evaluation	24	4
Activities not under evaluation (%)	14	NA

*Source:* MCC Impact Evaluation Database.

*Notes:* “NA” stands for Not Applicable.

Of the 57 experimental evaluations completed or still ongoing in fiscal year (FY) 2014–2015, twenty-three were RCTs (see Appendix D, Table D-1 for a list of MCC RCTs).<sup>42</sup> Twenty-three may not appear to be a large number of RCTs given the total number of evaluations. However, the percent of activities undergoing an RCT was about the same as the percent of activities not evaluated at all. When MCC’s experience with RCTs is compared to the World Bank’s, a major RCT proponent, we see that the MCC has initiated more RCTs relative to its size. When accounting for both organizations’ investment levels, the MCC appears to conduct at least twice as many RCTs as the World Bank. In 2015, MCC had annual investments of \$1 billion and executed 23 RCTs. The World Bank had annual investments of at least \$US 41 billion and executed an estimated 445 RCTs (Legovini, 2010; The World Bank, 2015).<sup>43</sup> Thus, for every billion dollars of investment, the World Bank executed 11 RCTs to MCC’s 23.

**Reframing Selection Bias.** There appears to be a discrepancy between what MCC invested in and most rigorously evaluated (see Table 6.2). More than half of MCC’s portfolio was in infrastructure, with one-third of all investments in transport alone. Yet, there was not a single RCT on transport. Most RCTs were in Agriculture and Rural Development. While Agriculture and Rural Development was a major focus at MCC early on and continues to play a role, the number of RCTs in Agriculture and Rural Development was disproportionate to investment in the sector. The same goes for the Health, Education & Community Services (HECS) sector. Education made up less than 5 percent of MCC’s total portfolio. This divide extended to the World Bank, whose RCTs likewise focused on education rather than transport—

---

<sup>42</sup> Because sometimes a randomized design might be converted to a quasi-experimental design or the evaluation may be abandoned altogether, the number of RCTs that MCC is running may fluctuate somewhat from year to year.

<sup>43</sup> The figure of 445 RCTs was arrived at based on an extrapolation from trends in the growth of impact evaluations and RCTs at the World Bank based on Legovini (2010).

21 percent of all impact evaluations were in education and 3 percent in transport while the sectors' investments made up 8 percent and 14 percent of all loans, respectively (Legovini, 2010).

Table 6.2

*Distribution of MCC Development Funding and RCTs by Sector and Year*

Sector	Average (%)	RCTs (%)
	2007–2013	2014–2015
Transportation	35	0
Agricultural & Rural Development*	21	57
Finance & Private Sector Development	6	9
Water Supply & Sanitation	9	4
Health, Education & Community Services	8	22
Governance	5	0
Energy	5	9
Total (excludes administration costs)	89	100

\*Includes irrigation and land tenure programs.

Source: MCC Annual Reports and MCC Impact Evaluation Database.

This distribution of RCTs across sectors was on par with the development community as a whole between 1981 and 2012 (see Table 6.3). Data from 3ie, an international NGO advocating for impact evaluation in development, illustrates the sectoral bias of RCTs. Most RCTs are in social sectors such as health (62%) and education (15%). The Water, Energy, and Transport sectors combined make up only 3 percent of RCTs.

Table 6.3

*Sectoral Distribution of RCTs in 3ie Database of Impact Evaluations*

Sector	Number of total RCTs	Percent of total RCTs (%)	Percent of all sector evaluations that are RCTs
Health, Nutrition & Population	1,228	62	83
Education	304	15	58
Social protection	139	7	41
Water & Sanitation	68	3	72
Finance & Microfinance	51	3	41
Information & Communications Technology	42	2	67
Agriculture and Rural Development	45	2	21
Private sector development	30	2	26
Public sector management	33	2	45
Environment & Disaster Management	23	1	30
Transportation	2	0	22
Urban development	3	0	16
Economic policy	0	0	0
Energy	0	0	0
Total	1,968	100	

*Source:* Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 8(1), 1–21.

A closer look at MCC's RCTs in the infrastructure sectors shows that the RCTs on electricity projects evaluated the impact of subsidized vouchers for households to connect to the existing electricity grid. The RCT in the water sector evaluated the impact of a small rural water project when MCC's investments were largely directed at upgraded and expanding utilities and urban water networks. None of the RCTs in the agriculture sector were on large-scale irrigation projects, which comprised the bulk of MCC's agriculture investments. When we include the projects most frequently evaluated with RCTs, farmer technical assistance and training (FTT), we see that over half of RCTs were directed at activities with some form of education or training.

MCC's RCTs reveal a deeper pattern that goes beyond sectoral bias—83 percent of interventions evaluated by an RCT are delivered at the individual or household level. The remaining 17 percent are delivered at the village or community level. It is not mere coincidence that the rise of RCTs has gone hand-in-hand with individualized interventions such as training, credit, and vouchers. Interventions directed at individual behavior to get at specific material effects (e.g., increases in income) are at the heart of randomized trials.

Development economists allude to this pattern but do not spend time explaining why individual interventions are favored. In his executive education training materials, Abhijit Banerjee makes it clear that it is “hard to evaluate monetary policy or freedom of the press” and that projects that are the most straightforward to evaluate “serve specific beneficiaries (individuals and communities)” (J-PAL, 2006). Barrett and Carter (2010) state that “[p]olitical economy questions that many believe to be of first-order importance in development are clearly not candidates for randomization. Nor are infrastructure issues or any other meso- or macro-scale intervention that cannot be replicated in large numbers” (Barrett & Carter, 2010: 527). Ultimately, there is greater fit between highly individualized, private interventions and the measureable impacts in an RCT. This is because the structural features of such projects make both randomization and the attribution of project-specific impacts more feasible.

Several MCC evaluation experts spoke of searching for activities where randomization would be most feasible. Despite the CGD Working Group's assertion that learning would drive design, it was first and foremost feasibility that determined if a project would be evaluated by an RCT. One MCC evaluation economist put it like this: “I think initially a lot of it was, ‘can we use an RCT?’ and if we can we should try. Like for the highway in [country] we couldn't use an RCT but I mean [country] had a lot of other sectors and we used RCTs in education and



agriculture” (Interviewee 3G, personal communication, June 29, 2016). This was sensible behavior on evaluators’ part. They were incentivized to evaluate programs and personally committed to excellence, which compelled them to use the supposed gold standard method. The issue was the method itself—no matter how much an evaluator wanted to, the method could not adapt to networked infrastructure such as highways, electric grids, or urban water systems because of sample size issues or spillovers. Consider this statement from an MCC evaluation economist:

Well the [projects] that work directly with people tend to be easier, so if you’re giving a scholarship, that’s in theory easier to randomize, if you’re training someone, if you’re working directly with a person...if you’re building a road, if you’re building water and sanitation, in theory you could randomize them if you have enough [sample], but often with infrastructure you don’t have the same numbers. I guess in my mind I think about feasibility (Interviewee 3G, personal communication, June 29, 2016).

The first practical issue in determining what to evaluate with an RCT is sample size. Where there may be several hundred classroom training sessions or scholarships in a particular project, there may only be a handful of roads or a single highway. Road construction and other infrastructure projects are costly, with some highway rehabilitation projects costing over \$US 200 million. One would need a sample of, at a minimum, several dozen highway-rehabilitation projects to, in theory, conduct an RCT. Meanwhile one could train thousands of farmers for one-tenth of the cost.

Another issue is oversubscription for project enrollment, which is shorthand for controlling exposure to the intervention. For an RCT to work the evaluator has to control when and how a unit of measurement, such as an individual or community, is exposed to an intervention. Such control is not feasible in networked infrastructure projects because they are directly antithetical to establishing and maintaining groups that do not receive the treatment. Another way to put it is to ask, “How does one control exposure to a highway”? Controlling

access to training, by contrast, is more feasible: “Compared to the energy and roads projects, RCTs are just amenable to any type of training program where you have a lot of people in a concentrated area and you can kind of say that these villages get it and these villages don’t, that kind of thing. It’s the oversubscription” (Interviewee 3H, personal communication, June 29, 2016).

A third issue is that large infrastructure projects, like national level policies, are public goods. RCTs are far more amenable to testing interventions based on private goods, even if they are publically provided, than public goods. An RCT’s need to establish a control group that cannot be exposed to an intervention for fear of “contamination” or spillover effects complicates its ability to experiment with public goods that are by design non-excludable and non-rival.<sup>44</sup> Thus, it is far easier to measure, for example, if an insecticide-treated bed net distribution program rather than a neighborhood storm-water drainage system lowers malaria rates. This is because of an assumption within randomized trials called SUTVA, or stable unit treatment value assumption.

Stable unit treatment value assumption is not something development economists like to talk about. You won’t find it anywhere in J-PAL’s executive training materials. A search of MCC’s hundreds of files on evaluation turned up the term, or its economic equivalent “partial equilibrium effect”, only once. The essential idea of SUTVA is that potential outcomes for any unit of an experiment are independent of the treatment assignment of any other unit under study (Sampson, 2010). In other words, any generation of internally valid estimates from an RCT is premised on the assumption of non-contamination or that those receiving the intervention

---

<sup>44</sup> Non-excludability refers to the extreme difficulty or cost required to exclude an individual from consuming a good or benefiting from it. Non-rivalry means that one person’s consumption of a good does not detract from or prevent another person’s consumption of that same good (Stiglitz, 2000).

(treatment group) had no social interaction with those not receiving the intervention (control group).

Any such interaction is called a spillover. According to two experts on randomized field experiments, spillovers are the Pandora's box of RCTs (Gerber & Green, 2012). To a certain extent, spillovers and externalities can be modeled and thus estimated, and academic economists have advocated as such (Duflo, Glennerster, & Kremer, 2007). But the selection of projects chosen for RCTs shows a desire to minimize spillovers. If spillovers are too frequent or too effective, it is analogous to expanding the treatment group and shrinking the control group, posing an existential threat to the experiment's validity. Great pains are taken to avoid spillovers, and by extension networked infrastructure projects because they are antithetical to establishing and maintaining groups that do not receive the treatment. Regarding the nature of infrastructure, Larkin (2013) states, "Infrastructures are built networks that facilitate the flow of goods, people, or ideas and allow for their exchange over space. As physical forms they shape the nature of a network ... Infrastructures are matter that enable the movement of other matter" (Larkin, 2013: 238).

A final point, and no less important, is that most of MCC's RCTs were for rural or anti-poverty community development programs. While this is largely because early Compacts were focused on rural development, it is also because experimenting on the rural poor and delaying or limiting poor farmers' exposure to interventions faced fewer barriers than, for example, limiting an important municipality's access to a rehabilitated road, water network, or power plant. The RCT experimental program is implicitly committed to addressing market failures by "nudging" risk-averse farmers to engage in "market learning" through subtle behavioral engineering (Berndt, 2015). Such motivations are not explicitly stated, but in the countries where MCC

works there are more poor individuals to “choose” from, thereby rendering programs aimed at large populations of poor people amenable to RCTs.

This bias towards highly individualized, private interventions that are excludable and rival is revealed in a popular sector for RCTs—Agriculture and Rural Development (ARD). Thirty-nine RCTs exclusively focusing on ARD could be identified in 3ie’s database of impact evaluations. Of those, 80 percent were focused on individualized, private interventions such as extension and training, credit, insurance, improved inputs, land titles, or cash transfers (see Table 6.4). Only 20 percent of the interventions were of a collectivized, public goods nature and even those were for small-scale, community-level interventions such as rainwater harvesting. Only one RCT was on governance. Not a single RCT was on irrigation.

Table 6.4

*RCTs Among Agriculture and Rural Development Projects in 3ie's Database*

ARD Project Type	Number of Projects Evaluated with an RCT	Percent of Total
Agricultural extension and training	11	28
Agricultural finance and insurance	10	26
Improved agricultural inputs	5	13
Agricultural information services	2	5
Rural cash transfers	1	3
Rural institutions	3	8
Rural infrastructure	2	5
Rural land reform	1	3
Seasonal migration	2	5
Other	2	5
Total	39	100

Source: 3ie (2015). *Impact Evaluation Repository*. Available: <http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/>

This pattern was repeated at the MCC. Of MCC’s 23 RCTs, fourteen are in Agriculture and Rural Development. Of these, every single one, except for an RCT in Burkina Faso on drip

irrigation, is focused on agricultural training and extension services, credit, rural business development, or land tenure. Burkina Faso's RCT also evaluates irrigation at the individual plot level rather than, for example, district-wide irrigation. Over 90 percent of MCC's Agriculture and Rural Development RCTs are on individualized, private interventions.

Established development economists argue that resistance to RCTs is due to “pure political interests” from sector specialists who are biased towards increasing funding for their sector's activities (Pritchett, 2002). However, while evaluation economists purport that RCTs represent the ideal approach for arriving at unbiased, apolitical estimates of project impact, RCTs present selection biases of their own. Not all projects are equally subjected to evaluation by RCTs. This should interest us for two reasons. If there is a bias towards which projects are evaluated with an RCT, then that subset of projects will either receive greater scrutiny from accountability-minded impact advocates or become the subjects of a specific “evidence discourse”. Other projects, meanwhile, will either escape scrutiny or have their “less rigorous” evaluations face questions of credibility now that the global development community has adopted new standards. Furthermore, as shown below, economic methods have political effects—they don't just observe, they intervene. The uneven application of economic methods means that their political effects are unequally experienced.

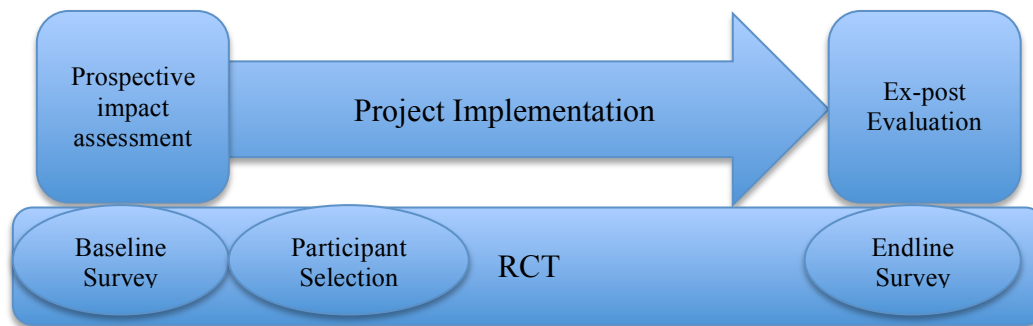
### **The Power of Method: Participant Selection**

RCTs were presented as research studies but in reality they were parallel projects that unfolded alongside a development intervention throughout implementation. They randomized project participants into treatment and control groups, and these groups needed to be maintained over the project's lifetime. In farmer training projects, choosing or targeting participants randomly conflicted with purposeful targeting based on the project's program logic. This logic

was geared toward maximizing crop productivity and income, necessitating project participants who were most likely to succeed at farming. The targeting criteria of project implementers were based on the diffusions of innovations paradigm and included a set of observable (e.g., assets, education level) and unobservable (e.g., motivation, skill) factors. The project implementers' use of unobservable farmer characteristics for project participation was particularly problematic for impact evaluation because it confounded implementer judgment with the training itself. To attribute impact to the project intervention—a key objective of accountability—MCC impact evaluators had to ensure that project participants were selected randomly.

**Conflicting Rationalities in Participant Selection.** RCTs are pitched and presented as research studies, but this is inaccurate. As **Figure 6.4** depicts, what an RCT really represents is a parallel project that unfolds alongside a development project. They are less like evaluations and more like interventions (Eyal and Levy, 2013). Other studies in development such as performance and ex-post evaluations take place *after* a project is complete. Social and environmental impact assessments take place *before* a project commences. An RCT, by contrast, is initiated before a project is fully designed and accompanies it *throughout* implementation. It therefore is continuously intervening in the project environment in a way that performance evaluations or impact assessments never do.

**Figure 6.4.** RCTs and Project Implementation



*Source:* Author.

RCTs intervene in projects in two fundamental ways: the placement of project participants into treatment and control groups and the maintenance of those groups' statuses over the life of the project. After deciding on an RCT the first step is random selection of participants. The rationality of randomization, however, is at odds with implementers' rationale for choosing program participants. This sets up the first act of resistance from implementers and requires the resolution of conflicting rationalities in evaluation economists' favor if an RCT is to survive. As mentioned earlier, evaluation economists argue that a randomly selected control group is necessary to avoid selection bias. Evaluation economists assumed that those responsible for implementing a project chose participants or participating locations, such as villages, based on politics or patronage. This is referred to varyingly as selection bias or "cherry picking". Economists also have methodological reasons for wanting to randomize. It allows for internal validity so impact can be attributed to a stylized project rather than a project's dynamic and detailed circumstances. Implementers meanwhile are trying to optimize project implementation to minimize cost. Some places and participants are easier to reach and have natural or infrastructural resources that facilitate project implementation and make outcomes more likely. They are also purposively choosing participants. This is because development projects are trying

to reach either one or more goals. For example, they may be trying to achieve an objective such as greater growth or relieve a condition such as water insecurity in a region or among a group of people. Therefore they almost always employ some aspect of *targeting* (Sen, 1992).

Targeting is usually based on a desire or need to maximize either efficiency or equity. This is derived from Okun's (1975) famous thesis in *Equality and Efficiency: The Big Tradeoff* where he states, "If both equality and efficiency are valued, and neither takes absolute priority over the other, then, in places where they conflict...some equality will be sacrificed for the sake of efficiency, and some efficiency for the sake of equality" (Okun, 1975: 88).<sup>45</sup> Regardless of whether implementers pursue efficiency or equality, the rationality around targeting project participants is decidedly *non-random*. Within this framework evaluation economists can be understood to be engaging in acts of *random targeting*.

When social control around RCTs was successful in Washington and the country capital, the RCT design limited implementers' targeting in two ways. One, the RCT prevented implementers from deviating from original observable screening criteria. Two, it prevented them from working with what they considered the most promising participants first. Both resulted in the rationality of randomization for project-attributable impact prevailing over the rationality of selecting participants to maximize project efficiency. Rational-legal authority was important for this effort in Washington, DC while more strategic forms of power were necessary to translate interests and extend the RCT network in country capitals and the field.

By taking a closer look at MCC's farmer technical assistance and training (FTT) projects, we can see how RCTs affected participant selection and project implementation. Among all

---

<sup>45</sup> Recent research has shown that efficiency may not come at the cost of equity and both goals can be pursued simultaneously (Andersen & Maibom, 2016; Basu & Stiglitz, 2016).



projects at MCC, FTT projects were most frequently evaluated with an RCT (see Appendix D, Table D-1). Among the 23 RCTs that had identified a control group, eight were FTT projects across the following countries: Armenia, El Salvador, Ghana, Honduras, Moldova, Morocco, Namibia, and Nicaragua. They comprised 35 percent of all RCTs initiated at MCC. Of these, five RCTs were completed (see Table 6.5), one is ongoing as of late 2016 (Namibia), and two were converted to quasi-experimental designs (Moldova and Morocco).

Table 6.5

*MCC's Completed FTT Projects with an RCT*

Country	Armenia	El Salvador	Ghana	Honduras	Nicaragua
Project	Water-to-Market	Production and Business Services	Commercial Training	Farmer Training & Development Assistance	Rural Business Development Services
Start – end year	2006 – 2011	2006 – 2012	2006 – 2011	2005 – 2010	2006 – 2011
Cost (millions)	\$18.2	\$64.4	\$62.2	\$26.6	\$8.5
Number of farmers trained	45,639	15,363	66,930	7,265	9,104
Implementer(s)	ACDI/VOCA	Chemonics	ACDI/VOCA, ADRA, IFDC	Fintrac	Chemonics, Technoserve
Evaluator	MPR	MPR	ISSER	NORC	UC Davis

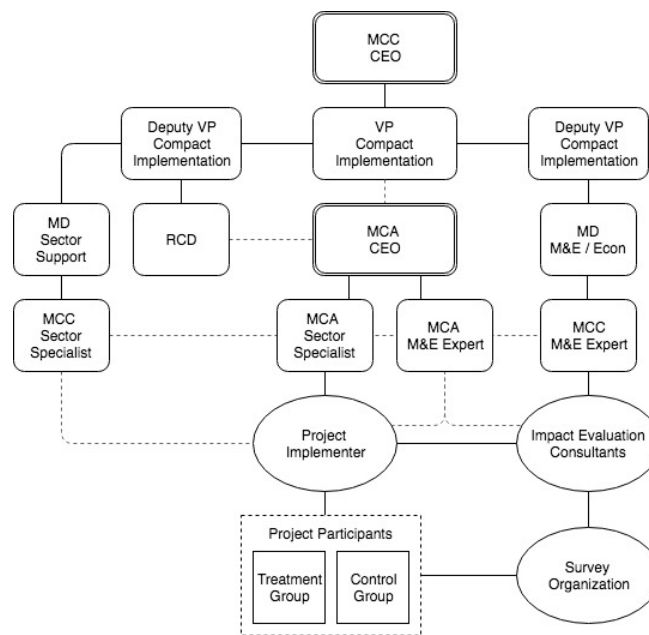
*Source:* MCC compact monitoring and evaluation plans and FTT project implementation documents.

*Notes:* ACDI/VOCA is a merger between a U.S. association of farmer cooperatives, Agricultural Cooperative Development International, and international development volunteer organization, Volunteers in Overseas Cooperative Assistance. ADRA is the Adventist Development and Relief Agency, a humanitarian and development agency of the Seventh-Day Adventist Church. IFDC is the International Fertilizer Development Center, an agricultural development agency with roots in the Tennessee Valley Authority. MPR is Mathematica Policy Research, a private policy research organization, NORC is the National Opinion Research Center at the University of Chicago, UC Davis is the economics department at the University of California at Davis, and ISSER is the Institute of Statistical, Social, and Economic Research at the University of Ghana.

The local government agency responsible for implementing the compact was called a “Millennium Challenge Account” (MCA). The MCA Chief Executive Officer and MCC Resident Country Director—the lead MCC representative in a country—would together oversee

compact implementation (see **Figure 6.5**). An FTT project implementer was usually contractually bound to the MCA and reported to an MCA official while an evaluation consultant was usually contractually bound and reported to an MCC evaluation economist at MCC headquarters in Washington, DC. MCC evaluation economists argued that this maintained a level of independence in the evaluation.

**Figure 6.5.** Implementation Structure of FTT Projects



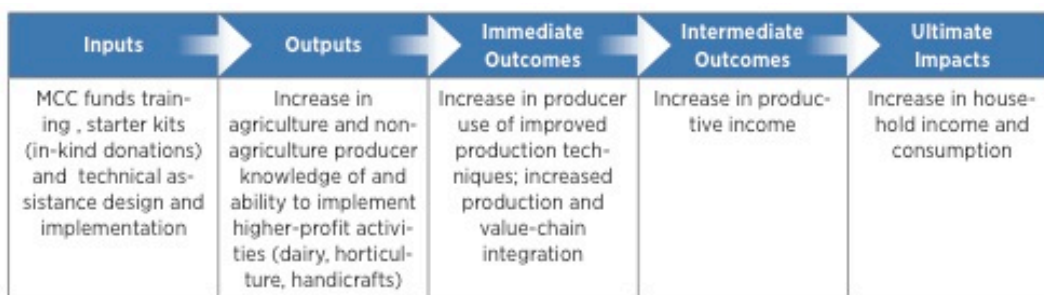
*Source:* Author

*Notes:* VP stands for “Vice President”. RCD stands for “Resident Country Director”. Solid lines represent formal relationships and reporting lines of authority while dashed lines represent informal relationships and collaboration. Rounded rectangles represent governmental entities while ovals represent non-governmental entities. Squares represent local populations. Not all impact evaluations and RCTs involve a separate Survey Organization. Sometimes the Impact Evaluation Consultants implemented surveys.

The remainder of this chapter will focus on RCTs’ effects on completed FTT projects. They were not only the most commonly evaluated with randomized trials but also illustrate why RCTs should be thought of more as political interventions than evaluations.

**Farmer Technical Assistance and Training Projects.** MCC’s FTT projects all operate on similar assumptions: 1) farmers’ lack of knowledge and skills is a key constraint for adoption of improved agricultural techniques, increased productivity, and crop marketing; 2) lack of inputs such as fertilizer and credit is a key constraint to increased productivity and investment; 3) increases in farmer knowledge will lead to increased business skills and improved access to markets; 4) when newly acquired knowledge on improved techniques is adopted and inputs applied, they lead to increases in farm productivity; and 5) increases in farm productivity lead to an increase in farm income that, in turn, leads to increases in overall household income. They also share similar program logic. Farmers receive training on planting, production, storage, and business management for non-subsistence crops and strategic inputs such as improved seeds and tools. Training and technical assistance translates to know-how and strategic inputs alleviate income constraints. Application of the training and inputs leads to higher productivity of non-subsistence crops and value chain integration. This in turn leads to higher income. **Figure 6.6** is an example of a simplified program logic from the El Salvador FTT.

**Figure 6.6.** A Simplified Program Logic Framework from the El Salvador FTT

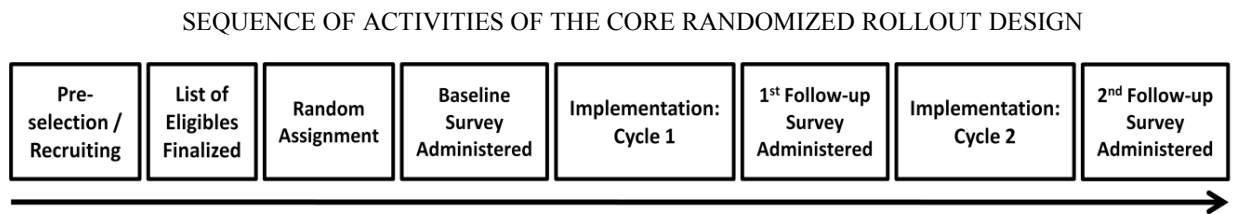


*Source: MCC (2012c). Measuring Results of the El Salvador Production and Business Services Investments. Washington, DC: Millennium Challenge Corporation.*

Each of the projects used what is called a randomized rollout design because it was controversial to entirely deny farmers services as part of a government program (see **Figure 6.7**).

The randomized rollout is one approach J-PAL advocates for when pure randomization is not possible (Glennerster & Takavarasha, 2013). Unlike in a “pure” RCT where a control group is established and then denied services, in a randomized rollout some farmers were randomly sorted into an early treatment group while others were placed in a late treatment group. The late treatment group would have to wait anywhere between 12 and 36 months to receive services. Impacts on the early treatment group would be observed during this time frame.

**Figure 6.7.** Randomized Rollout Design for FTT Projects



*Source:* Campuzano, L. & Blair, R. (2011). *Final Impact Evaluation Design for the Production and Business Services Activity of the Productive Development Project – Revised* (ESVED-264). Princeton, NJ: Mathematica Policy Research.

**Targeting in Farmer Training Projects.** The program logic of MCC FTT projects was aimed towards efficiency. The projects’ objective was increased farmer productivity through behavior change in farmers’ relationship with improved techniques and inputs. Implementers were usually responsible for meeting benchmarks for increases in farmer adoption of improved inputs, crop productivity, and farm incomes. The logic behind efficiency-based targeting in farmer training is rooted in a concept derived from rural sociology. In the 1930s Bruce Ryan and Neal Gross set out to understand how hybrid corn seed diffused in two Iowa towns. Their research resulted in a program on “diffusion of innovations” that transformed not only agricultural research but also the way behavior change was studied in the social sciences. A key takeaway from this research program relevant to targeting is how an agricultural innovation, such

as improved inputs, spreads among farmers. Ryan and Gross (1943) grouped farmers into five categories: innovators, early adopters, early majority, late majority, and laggards. For an innovation to be adopted quickly and reach “critical mass”, it needed to be accepted and incorporated by a community’s early adopters. Typically, early adopters are opinion leaders and held in high regard by neighbors and other community members. They have higher social status, educational attainment, and financial liquidity than their peers. They are also more cosmopolitan (Rogers, 1962).

The diffusion of innovations research program resulted in a major paradigm shift in agricultural research in the 1940s and the main tenets of the thesis are still accepted today (Valente & Rogers, 1995). Thus, implementers believe that the more quickly and widely improved techniques and inputs are adopted, the more quickly and widely these constraints to farmers’ productivity and income could be removed. Key to this strategy is the kind of farmers that are targeted and when they are trained. An implementer of Ghana’s FTT put it thus:

If you are introducing a new technology that is maybe very new to the farmers, you always want to make sure that you look for those who are the most innovative people, the innovators and the early adopters...if they pick it up and the thing works you find that it’s much, much easier...to see higher levels of adoption even among the late adopters...But if you start with those who are very skeptical and you introduce to them the new technologies they might not do it, or not do it well, then it presents a poor picture of the new technology (Interviewee 3N, personal communication, August 1, 2016).

In the absence of randomization or random targeting, FTT project implementers, interested in maximizing project efficiency targeted participants on seven main criteria: i) membership in a farmer based group; ii) education; iii) innovativeness; iv) ability to disseminate ideas; v) resources; vi) wealth, and vii) social standing (Phillips, Waddington, & White, 2014). These are a mix of binary variables (e.g., membership) and continuous variables (e.g., education, wealth). They are also a mix of “observable” (e.g., resources) and “unobservable” (e.g.,

innovativeness, social standing) characteristics. The most popular criteria were membership in an organized farmer group or cooperative and basic literacy and numeracy skills. Practical criteria are also important and implementers look to target farmers whose locations are accessible, are available to fully participate in training, and are interested in the training and its objectives (Phillips et al., 2014). These are employed to limit the waste of operational resources.

The same review of farmer training eligibility criteria also summarized how farmer-training projects targeted participants. Data from 58 projects showed that 80 percent of projects targeted participants through categorical targeting. Categorical targeting is a method that determines eligibility based on whether an individual belongs to a specific category. This is also referred to as group targeting. The categories employed, such as age, gender, or ethnicity, are usually fairly easy to observe, hard to falsely manipulate, and are relevant to the program's objectives (Coady, Grosh, & Hoddinott, 2004). Typically individual or household assessments would determine who ultimately participated in the projects (Phillips et al., 2014). Field level technicians (i.e., field bureaucrats) conduct individual or household assessments to determine if potential participants meet eligibility criteria. In farmer training programs, assessment was almost always in the form of community- or implementer-based selection of participants, rather than a means test (Phillips et al., 2014).

MCC FTT projects employed similar targeting criteria and mechanisms to other farmer training projects. Table 6.6 shows the inclusion criteria used for farmer selection in MCC's Central American FTT projects.

Table 6.6

*Eligibility Criteria for Participation in MCC FTT Projects*

Nicaragua	Honduras	El Salvador
Area of land holdings	Area of land holdings	Access to land
Access to a water source	Access to a water source	For livestock, minimum number of cattle
Possession of land title	Farm that does not flood and is accessible by a paved road	Desire to participate in project activities
Minimum area of land cultivated in past	Not an existing user of new technologies	Experience or interest in producing and selling horticulture
Minimum age	Willing to identify and organize neighboring farmers	Interested in forming alliances with other producer groups
	Commits to implementing improved technologies and practices	Interested in marketing and selling products with other groups
	Promises to invest the necessary time and money for success	Willingness to co-finance farm investments

*Source:* Official MCC FTT implementer project inception and interim reports.

MCC FTTs had a mix of binary and continuous variables and observable and unobservable characteristics. The most common criterion was access to a certain area of land holdings. This made immediate sense since without land farmers could not apply and practice new techniques. Other criteria included a willingness to collaborate with others and provide a contribution to on-farm investment. Implementers believed both criteria signaled a commitment to improved agricultural productivity and sales.

While implementers may apply a set of categorical screening criteria to narrow their target group they also have good reason to allow their field technicians to exercise discretion and deviate from or refine that criteria. Implementers may choose to do this for two reasons. First, unobservable criteria used for the first step of categorical targeting do not render themselves legible easily. How does one measure “commits to implementing” or “interested in marketing”?

Second, even when legible, some of the criteria are based on binary characteristics while others are based on continuous characteristics. While “possession of land title” is a binary characteristic, “area of landholdings” is continuous (and even terms like “possession” and “title” are complex and open to interpretation). If landholdings are a desirable trait for potential program participants, implementers may consider those with more land as more eligible. Since FTT implementers are trying to employ efficiency-based targeting to achieve behavior change and innovation diffusion goals, they have an incentive to select participants who are best positioned to succeed through training. These may not be the farmers that were originally selected for treatment based solely on observable criteria or around the margins of a cutoff. Subjective assessments of both continuous variables and unobservable characteristics may result in a different set of farmers appearing as better suited for project goals after “ground truthing” or field technicians’ verification of specific farmers’ suitability for project participation.

**The Political Effects of RCTs: Participant Selection.** A closer look at the Honduras FTT will illustrate how an implementer goes about adjusting participant selection without the restriction of an RCT. While an RCT was planned for the Honduras project, lack of early involvement of the evaluator, lack of contractual authority to adhere to a randomized design, and tepid support from the MCC Resident Country Mission limited social control. As a result, the implementer in Honduras, Fintrac, a 25-year-old Washington, DC based company that “develops agricultural solutions to end hunger and poverty”, could deviate from the originally selected treatment group.<sup>46</sup> The evaluator, the National Opinion Research Center (NORC), selected treatment and control villages based on geographic information systems data on elevation, soil capacity, rainfall, vegetation, distance to nearest major river, and other variables. NORC

---

<sup>46</sup> Further information on Fintrac can be found at: <https://www.devex.com/organizations/fintrac-6761>



matched pairs of villages that looked similar and drew treatment and control samples from each. Farmers could then be selected for early and late treatment from the respective villages based on observable screening criteria provided by Fintrac (NORC, 2013). These selection criteria are in Table 6.6.

NORC conducted farmer selection based on criteria provided by Fintrac.<sup>47</sup> While NORC selected 1,017 farmers in 113 villages for early treatment, Fintrac chose to work with only 28 of these farmers across 19 villages (NORC, 2013). The rejection of so many treatment group farmers meant that entire villages selected for treatment dropped out of the study and the randomized design collapsed. The evaluator concluded that basing farmer selection on observable Fintrac criteria was “counterproductive”:

It was clear that the original design concept of stratifying farmers based on the set of quantitative criteria provided to us by Fintrac was counterproductive...we concluded that we could not replicate Fintrac’s selection process for two reasons: (1) it contains elements/criteria that could not be quantified and depended on some element of subjective assessment by the Fintrac Field Technician of a farmer’s motivation, ability to learn and grow, and willingness to follow program requirements; and (2) the selection criteria evolved over time, based on lessons learned during implementation (NORC, 2013: 16).

Following the rejection of NORC’s treatment group, NORC conducted interviews with Fintrac field technicians to discover why their selection of treatment farmers had deviated so far from their own. They discovered that Fintrac field technicians replaced NORC’s originally selected farmers with farmers who were either more motivated and/or more aware of the level of effort required to succeed in the program. Some farmers reached out to Fintrac field technicians to request participation while the field technicians identified other farmers who had already constructed raised beds and other pre-planting steps. To field technicians, these were both signals

---

<sup>47</sup> Fintrac did not want to enter communities to select farmers only to turn around and tell farmers that they were going to postpone service delivery for 18 months. For similar reasons, this practice was common across several MCC FTTs.

of program commitment and potential program success (NORC, 2011). According to an MCC agricultural specialist:

My perspective is that not all farmers indeed turn into commercially oriented farmers. Let's just start from that basic fact. To assume that any and every farmer will be a good candidate and so let's just randomize, it just doesn't make sense because, again, not every farmer has the capacity to be commercially oriented and there are a lot of other obstacles ... transforming yourself from being a corn farmer into an irrigated horticulture farmer, the costs and the level of knowledge [required], it's a big step (Interviewee 3L, personal communication, July 8, 2016).

According to the field technicians, selecting farmers based on their motivation and completion of pre-planting steps raised the quality of participants in the program, reduced drop out rates, and allowed field technicians to focus more on technical services rather than recruitment of farmers who ultimately may be less committed and more likely to quit (NORC, 2011). The field technicians may have also been able to tacitly identify the most capable farmers that would thrive in the program. The lead evaluator for the Nicaragua farmer training RCT, an economics professor at The University of California Davis, and several of his colleagues attempted to understand what they called “impact heterogeneity” in the Nicaragua FTT project. There were “high performing” participants and “low performing” participants among eligible and randomly assigned farmers. They said that this “raises the question of what distinguishes the people who benefit a lot from a program from people who didn't” (Carter, 2016). They tried to answer this question by examining what they called technical efficiency—an estimation of farmers' cultivated land area and total farm labor applied as a proxy for unobservable farmer capacity—had on farmer participation and income.

They analyzed the evaluation data and found that variation in farm incomes among program participants was largely related to something that's different between high and low performing farmers, but it was not necessarily farmers' technical efficiency. Farm incomes did

not rise as farmers' technical efficiency rose. According to the researchers, "So what this suggests is that there are other things going on, it's not just about technical efficiency, there have got to be other factors driving impact heterogeneity...Clearly other unobservables [are] in play" (Boucher, Malacarne, & Carter, 2016: 35). As one MCC evaluation economist who was familiar with the exercise put it:

In Nicaragua they found a little bit of an impact, but the interesting thing that goes to the implementer's point...some people did really well and some people didn't do well at all...But it was not related to anything observable...there's something going on where people are winners and losers in a sense, and in Honduras that's what they wanted to pick [participants] on, right? In a way they didn't just want to randomize (Interviewee 3E, personal communication, June 24, 2016).

Despite field technicians' potential to identify the most capable farmers, evaluation economists do not want implementers to deviate from the random assignment criteria used during implementation or select based on unobservable characteristics because doing so would confound program effectiveness and eligibility. The selection of participants based on field technicians' subjective criteria rather than the randomization process introduces either the field technician's judgment, farmers' self selection, or the relationship between field technicians and farmers as a determinant of project success rather than the training, improved techniques, or inputs themselves. This undermines an RCT's attempts at internal validity because judgment and relationship quality are difficult or impossible to measure and external validity because these project characteristics cannot be scaled up or replicated in other places the way standardized training, techniques, and inputs can.

Even while implementers and field technicians appreciated the ability to deviate from observable screening criteria, some observable characteristics are known to facilitate the adoption and spread of improved practices. Educational attainment, wealth, and cosmopolitanism were all important determinants in the adoption and diffusion of improved

inputs. A review of farmer training programs in developing countries showed that farmers' years of education was a statistically significant determinant of adoption of practices, improved yields, and diffusion of lessons learned to neighbors (Phillips et al., 2014). Implementers of Ghana's FTT echoed this sentiment: "Even among the farmers, it is very clear even here in Ghana and in many places that you find that literate farmers, they tend to be higher adopters of technologies. It's easily seen here if you take basic things like adoption of herbicides, some insecticides, new varieties [of seed]; you always find literate people going for it first" (Interviewee 3N, personal communication, August 1, 2016). In Ghana, these more capable farmers are known as "nucleus farmers". Nucleus farmers are important "hubs" in farmer and agricultural production networks. They are socially well connected, have important agricultural assets, and knowledge of agricultural value chains and markets. By contrast, "out-grower" farmers have relatively fewer assets, including smaller land holdings, and limited know-how of value chains and markets. As a result, each out-grower farmer is partnered with a nucleus farmer that provides technical support and facilitates market access (Lucas, 2006a).

Rather than allowing subjective assessments of unobservable characteristics to distinguish early project participants from late or non-participants, evaluation economists "balance" the early and late treatment groups so that they are *indistinguishable* based on *observable* characteristics. In MCC FTT projects, usually over a dozen observable characteristics of demographics, resources, and farmer practices were assessed for statistically *insignificant* differences. This facilitates impact attribution to a stylized project rather than any differences between the two groups. In other words, the first farmer trained should look very similar to the last farmer trained on predetermined observable criteria such as age, educational levels, and wealth.

Contractual authority and MCC Resident Country Mission consent permitted evaluation economists to exclude implementers' and field technicians' subjective assessments. To build support for early and late treatment groups, however, evaluation economists also needed support from those closest to participants such as community leaders and local officials that might share implementers' and field technicians' rationalities. Evaluators usually staged a participatory randomization ceremony (or multiple ceremonies) to build local support for randomization. A World Bank economist conducting an RCT in Cameroon described how randomization ceremonies—public events where potential project participants were randomly sorted into treatment or control groups using material devices such as software programs and ping pong balls—appeal to fairness and transparency. These moral appeals can help justify the “politically difficult” task of delaying benefits to the control group and convincing participants’ patrons, such as local leaders, that randomization was in everyone’s best interest:

Our experience in many countries is that public randomization ceremonies are an excellent platform to build support for randomization and for the entire impact evaluation process. In Cameroon, we organized public randomization ceremonies in three Regions to assign health facilities to four study groups in an impact evaluation of performance-based financing (PBF) in the health sector... This made the randomization process completely fair and transparent to all health facilities participating in the study (De Walque, 2013).

Once early and late treatment groups were established, the implementers, which were contractually bounded to the MCA, had little choice of whom to train and where to train them. They had to stay within the confines of the original, codified eligibility criteria and work with farmers in the order assigned to them. Unlike Honduras, the remaining four MCC FTT projects’ adhered to the assignment of treatment and control groups and, contrary to the rationality of efficiency-based targeting, the two groups looked statistically alike (see Table 6.7).

Table 6.7

*Key Characteristics Among Early and Late Treatment Groups: MCC-FTT Projects*

	Armenia		Nicaragua		El Salvador		Ghana	
	T	C	T	C	T	C	T	C
<b>Age (years)</b>	55	55	50	52	45	44	45	45
<b>Education</b>								
More than secondary school education (%)	15	16						
Education (years)			4.3	4.9				
“More than basic” education (%)					9	7		
At least secondary school education (%)							11	10
<b>Wealth</b>								
Land owned or rented (hectares)	1.8	1.8	25.9	28.8			1.3	1.2
Total net annual income (US \$)					1,577	1,700		

*Source:* Projects’ respective evaluation final reports.

*Notes:* “T” stands for treatment group; “C” stands for control group. Control group is the “late treatment” group. All results between treatment and control groups are statistically insignificant. Honduras was excluded from the table because the integrity of the control group was compromised.

### **The Power of Method: Intervening in Project Execution**

Randomizing participants was only the first step; the control group required maintenance over the project’s lifetime. This meant limiting contamination, or contact, between treatment and control group participants, especially early in the project. Units such as farmers, farmer organizations, or farming communities were sampled to maintain spatial and temporal distance. This also presented tensions with the implementers’ rationale: implementers wanted to encourage contact between “pioneers” and other farmers to boost project outcomes through

improved innovation diffusion and train everyone as soon as possible rather than waiting to train the control group. To keep implementers from training control group farmers sooner rather than later, the evaluation team had to continuously negotiate with the implementers and persuade them to maintain the control group, which was more feasible through an in-country presence.

**Extending the RCT Rails.** Once conflicting rationalities around participant selection were resolved in evaluation economists’ favor, the next task was to maintain the control group over the life of the project. There needed to be a treatment group that received the intervention and a control group that was denied the intervention (altogether, for a period of time like in FTT randomized rollout evaluations, or not encouraged to participate). These two groups needed to be surveyed in two distinct temporal periods: baseline (t1) and endline (t2). The four observations needed for each unit of analysis is known as the difference-in-differences matrix. The difference-in-differences estimator, or project impact, is the difference between the treatment group and the control group and over time (t2-t1) (see Table 6.8):

Table 6.8

*Differences-in-Differences Matrix*

	Before Change	After Change	Difference
Treatment Group	Yt1	Yt2	$\Delta Y_t = Y_{t2} - Y_{t1}$
Control Group	Yc1	Yc2	$\Delta Y_c = Y_{c2} - Y_{c1}$
Difference			$\Delta\Delta Y$ $\Delta Y_t - \Delta Y_c$

Source: Author.

Notes: “Y” stands for year, “t” for treatment, “c” for control, and “Δ” for “change in”.

In order for RCTs to succeed, the difference-in-differences matrix needed to be recreated in the project implementation landscape. Project impact estimates need a conducive

environment or experimental apparatus in which economic facts can survive. Just as Louis Pasteur's vaccine would not work across France unless his laboratory conditions at Pouilly le Fort were first extended across rural France, valid estimates of development programs or projects' impacts cannot be established unless certain field conditions are first met. Using the metaphor of a railway, Latour (1983) says "you cannot drive a locomotive through a field" (Latour, 1983: 155). He meant that for the trains or scientific facts to work, scientists must first "extend the rails" (Petty & Heimer, 2011) or sufficiently recreate their laboratory conditions in the outside world so that their facts can survive.<sup>48</sup> While evaluation economists did not have a laboratory *per se*, the difference-in-differences matrix is a micro, calculative world whose rails had to be extended to the macro, implementation landscape.

The FTTs sampled units in a way to maximize the temporal and spatial distance between the treatment and control groups. This was to observe impact as well as guard against "contamination". MCC's evaluators for the Armenia FTT succinctly described contamination and how to prevent it. The evaluators' description is worth quoting at length:

Farmers who are in the control group of village clusters would not be offered water management training until several years later. All villages will ultimately be provided training, and random assignment is used to determine *when* they are offered training. Randomly assigning entire village clusters in this way, rather than individual farmers or villages, guards against contamination of the control group....There are two types of contamination. The first type of contamination is if farmers in control group villages nonetheless participate in training. This could be problematic if control group members hear about the training activities and show up to training themselves. A different kind of contamination could occur if farmers who participate in training teach farmers in the control about the techniques they learned. Either of these types of contamination would be problematic for the evaluation...Generally, [the implementer] has chosen village clusters that are sufficiently far apart geographically to ensure that there is little chance that farmers in a control group village cluster would either participate in the training or learn about water management techniques through other means....While we cannot completely eliminate the possibility of contamination here, it will be important for the

---

<sup>48</sup> I thank Margarita Rayzberg for introducing me to the work of Petty and Heimer.



planned implementation to strive to avoid such contamination problems (Fortson and Rangarajan, 2008: 4-5).

Avoiding contamination limited both the project's reach and knowledge diffusion objectives. A World Bank economist who conducts randomized evaluations of training programs described why implementers are against limiting reach: "There can be a huge temptation for the people delivering the program to dip into the control group. This can either [be] because they have been contracted to train or deliver the program to a certain number of people, so that every empty seat directly costs them, or just because they see they have space and genuinely want to help as many people as possible" (McKenzie, 2015). For example, when it comes to farmers training other farmers (i.e., diffusion), an MCC agricultural development specialist said "[t]hat's a good thing, it's not a proper evaluation, but that's a good thing" (Interviewee 3J, personal communication, July 6, 2016). The rationale for limiting contamination was in conflict with the rationale of implementing to maximize reach, diffusion, and efficiency. Extending the RCT rails had implications for project implementation and therefore met resistance.

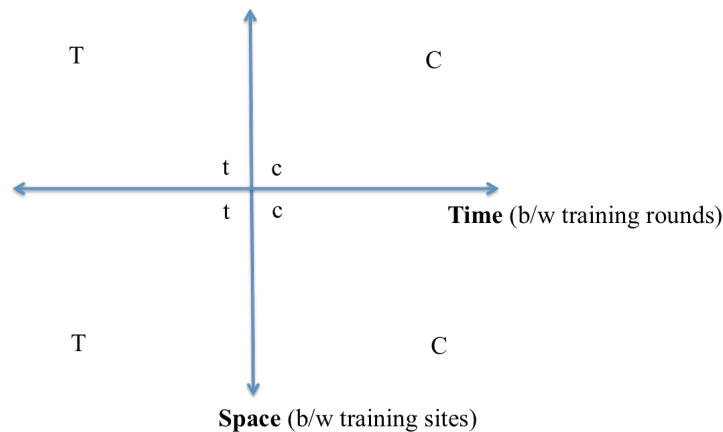
Resistance to RCTs was a topic of discussion at a joint MCC USAID Roundtable on *Agriculture Impact Evaluations: Learning What Works* at MCC headquarters on July 21, 2011. MCC and the United States Agency for International Development (USAID) brought together MCC evaluation economists and staff, MCC's evaluation contractors, and MCC agriculture sector specialists to "[i]dentify challenges, lessons learned and successful strategies for implementation of rigorous impact evaluations in the context of agriculture and rural development programs" (MCC, 2011: 1). The Roundtable consisted of panels and breakout sessions to "foster discussion between agriculture project implementers, impact evaluators, and donor agencies" to describe challenges and propose solutions to rigorous evaluation of

agriculture and rural development projects (MCC, 2011). Among the reasons identified for resistance were RCTs' implications for project scope, design, and flexibility. Regarding project scope, participants cited that: “[r]equiring a control group that is comparable to the target population can reduce the overall reach and scope of programs” (MCC and USAID, 2011: 3). Regarding implications for project design, they stated: “Agricultural development practitioners argue for an implementation design that maximizes the reach of the intervention...they may also seek opportunities to build on spillover effects. However, both of these approaches may contaminate the control group” (MCC and USAID, 2011: 3). Finally, regarding project flexibility, the report stated that “[a]dherence to an impact evaluation methodology may limit implementers' ability to adapt implementation approaches” (MCC and USAID, 2011: 3).

According to an MCC agriculture specialist, without an RCT “some things would have been done considerably differently” (Interviewee 3J, personal communication, July 6, 2016) (see **Figure 6.8**). Implementers would have trained everyone sooner on a continual basis rather than leaving a gap of at least one year to observe impacts on early treatment groups: “It would have been optimal to train all the farmers as fast as you could which we couldn't because we had to stagger the training. We would have done [the training] on a faster rolling basis” (Interviewee 3J, personal communication, July 6, 2016). Second, they would have started training and services in an area where they could build on the existing physical, natural, and institutional infrastructure left behind by similar projects in the past. From there they would have gone on to train sequentially in nearby locales taking into account accessibility such as road network quality. For example, another MCC agricultural specialist said, “[the implementers] went to this one town that was technically in the municipality that we were going to do, but the road to get to the town is really bad, so one of our criteria is that they really have to have year round access to

markets, road wise, and this road, it's so bad that during the rainy season you cannot access [the town]" (Interviewee 3L, personal communication, July 8, 2016). RCT designs did not take into account such conditions.

**Figure 6.8.** Schema of Training Approach With and Without a Randomized Rollout Evaluation



*Source:* Author

*Notes:* Letters represent trainings conducted by implementers. Lowercase letters represent implementers' approach without a randomized rollout design. Uppercase letters represent implementers' requirement with a randomized rollout design.

**The Political Effects of RCTs: Implementation.** The implications of successfully “extending the RCT rails” can be seen in the Ghana FTT where the RCT’s control group was maintained throughout the project’s lifetime. According to the joint MCC USAID Roundtable’s participants, the lesson from successful impact evaluations was that “[e]arly engagement...fosters deeper understanding of the implementation approach as well as the impact evaluation design methodology. Impact evaluations go more smoothly when evaluators involve implementers in key aspects of methodology design and continue this engagement throughout implementation” (MCC and USAID, 2011: 8). Critical to this engagement was ongoing

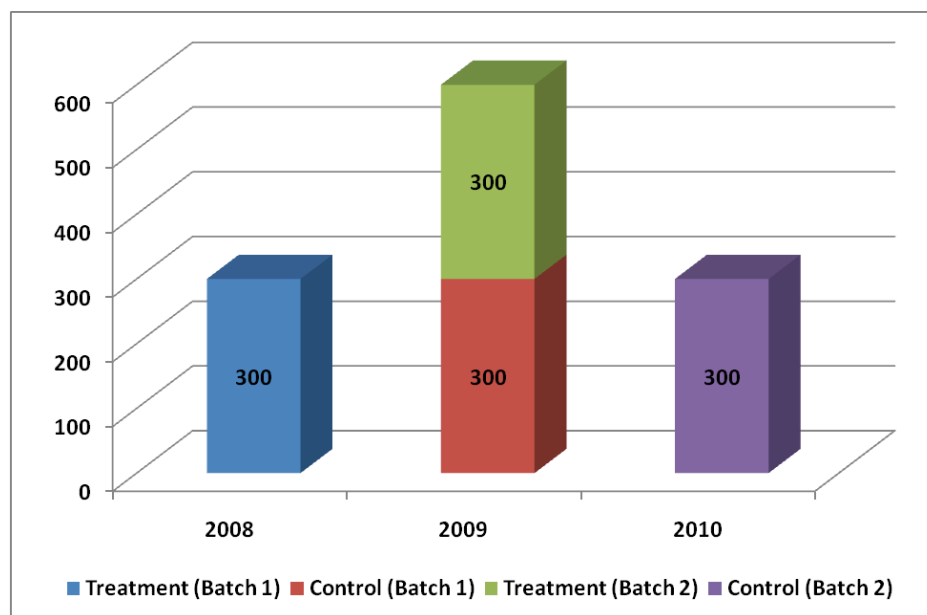
communication: “The integrity of the project implementation as well as the impact evaluation design will benefit from ongoing communication...All participants should be encouraged to raise concerns early and often, so that solutions can be found that best manage tradeoffs between implementation flexibility and evaluation validity (MCC and USAID, 2011: 9). Evaluators and evaluation consultants’ presence in the field facilitated communication. The Roundtable’s report stated that “[i]t helps for the impact evaluation team to have representation in the country—not only a full-time staff who can respond rapidly to changes in the project, but also team leaders who can visit the project sites, understand the implementation approach firsthand, and communicate directly with implementers about their challenges and concerns” (MCC and USAID, 2011: 9). In Ghana, like other successfully executed RCTs, engaged implementers and ongoing communication were critical to maintaining the control group and reshaping implementation.

The Ghana FTT delivered approximately 30 hours of agronomic, business, and marketing training to farmers. Participants also received a starter pack and the option to access credit. The starter pack was valued at approximately \$230 and included fertilizer, one acre’s worth of seeds, boots, a facemask, gloves, and some money for land clearing (Adventist Development and Relief Association, 2011). The project was implemented across 30 districts in three “zones” of the country: the Northern Zone, Afram Basin (“Afram”), and Southern Horticultural Belt. Each zone had a separate implementer, known as a Regional Implementation Consultant. The project had one evaluator, the University of Ghana’s Institute for Statistical, Social, and Economic Research (ISSER), which is located outside of the capital, Accra. Unlike in Honduras, the Ghana FTT control group was maintained during project implementation.

In Ghana, the random selection was done at the Farmer Based Organization (FBO) rather

than individual farmer level. Farmer based organizations either had close to 50 members or were encouraged to join forces with others or split up to reach 50. Once an FBO was selected, every member was eligible for training, technical assistance, and the starter pack. The program aimed to train 60,000 farmers in 1,200 FBOs (the final number of farmers trained was higher at 66,930). The FBOs received training in batches to stagger the training over time. Each batch consisted of 600 FBOs and was split into a treatment (early treatment) and control (late treatment) group. The time in between trainings for each respective batch's treatment and control groups was one year. However, the time in between the treatment group for Batch 1 and control group for Batch 2 was two years. These gaps in training were necessary to observe project impact on early treatment groups (see **Figure 6.9**).

**Figure 6.9.** Training Rollout in Ghana's FTT



Source: ISSER (n.d.). *Approach to Evaluating Impact of the training component of the MiDA Programme*. Accra: Institute of Statistical, Social and Economic Research.

Farmer Based Organizations were randomized according to stratified cluster sampling. Here stratified refers to the stratum, or zonal region, within which randomization occurred. Within these strata, FBOs were randomized into either Batch 1 or 2 and then either treatment or control. This was done to minimize the effect of either an FBO's regional location or technological sophistication on project impact estimates. Here "cluster" refers to sampling by FBO (i.e., 50 farmers) rather than by farmer. This was done to account for positive externalities (i.e., spillovers) within the FBO cluster. By measuring impacts at the FBO level, one captures the spillover effect on untrained farmers in the cluster (while all members were eligible for training, not all were required to participate). This is a strategy that J-PAL advocated for in its executive education to account for positive externalities through a case study of deworming in Kenya (Poverty Action Lab, 2005). This does not resolve the issue of spillovers across clusters, however.

Once sector specialists and MCA senior officials were in agreement over random selection, implementers were bound by their contracts to sequence training in a certain way. The Millennium Development Authority (MiDA), the local MCA office responsible for managing the compact, contracted the implementation consultants to manage the FTT and ISSER to conduct the evaluation. Implementation consultants, in turn, contracted technical training service providers to deliver technical assistance and training to farmers. Prior to the FTT's start, ISSER generated farmer lists for training based on training batches and control or treatment status. MiDA would then approve "task orders" for the service providers, which the implementation consultants would oversee and manage. Task orders governed the contractual relationship between MiDA, implementation consultants, and service providers. Over 200 task orders were issued during the project. When asked why the implementer did not train the control group

earlier given its incentives to do so, one implementation consultant manager said, “[b]ecause before we can do the training we have to get approval from MiDA” (Interviewee 3N, personal communication, August 1, 2016). The service providers would also have to keep photos and written records of farmer attendance and services rendered, which implementation consultants would then send to MiDA in Accra for verification.

But rational-legal contractual authority could be undermined in informal ways, even if only partially. One MCC agricultural specialist remarked,

The biggest challenge...is the implementers don't want to do [the RCT]. They don't want to have a control group, they don't want to have any reduction in flexibility in how they implement their project. So if you say, these X communities, you cannot touch them, [they say] ‘well, how am I gonna get my results?’ So we have a lot of resistance, and we've moved towards putting out [Requests for Proposals] that don't give them an option, you have to [have a control group], and even so they chip away at it (Interviewee 3I, personal communication, June 30, 2016).

Formal contractual authority was insufficient for limiting the erosion of the control group over time and also for managing local leaders' and participants' resistance. Local leaders, such as district leaders, village chiefs, and executives of rural cooperatives, were also often opposed to a control group. They advocated on behalf of their respective constituents. When their constituents were denied or delayed services, these leaders were not happy and made it known. This is in part because potential project participants were unhappy with being in a control group. They too had grievances and opposed having services denied or delayed.

Given this resistance, maintaining the control group required persuasive power in the form of “constant negotiation” that was “tedious work” (Interviewee 3E, personal communication, June 24, 2016). This work was tedious because, unlike MCC sector specialists and MCC and MCA senior officials, these actors' interests had not been translated because it was more difficult to enroll and mobilize these dispersed local actors into the RCT network.

Implementers were not invited to early MCC workshops on impact evaluation, largely because they had not yet been contracted. But even if they had, the question of whether impact could be as easily framed as learning rather than accountability remained. Agriculture and rural development project implementers have their own conceptions of learning and experimentation that are rooted in learning as doing, or praxis. For evaluation economists project participants such as farmers were an afterthought. While the number of participants, which were in the thousands, precluded their participation in research design and planning, the RCT rationale nonetheless ignores participants' agency. The RCT considers participants as vessels into which projects are poured and outcome data later retrieved (Breslau, 1998). The inability to deepen the network among participants meant that evaluators had to persuade those involved in project implementation such as MCA management and in-country MCC leadership but particularly the project implementer.

An evaluator's in-country presence made this persuasion all the more feasible and effective. In its evaluation work plan, ISSER said, "Evaluation of the FBO training and irrigation components relies on randomization of the entry of FBOs into the program. In each of these cases, measuring the impact of the program will require intensive, frequent, timely and close cooperation and exchange of information" (ISSER, 2007: 27). In other FTTs where the control group was maintained, the evaluator was not locally situated. Close cooperation in this case meant as needed conference calls that could be as frequent as every week between MCC, MCA, evaluators, and sometimes implementers; trips to the country capital on a quarterly basis by evaluation economists and/or MCC evaluation consultants; and occasional site visits to training locations. An MCC evaluation economist spoke of the types of persuasion employed: "I think the hardest part of our job is the negotiation and just making sure that people understand



and value what you're doing...we would talk and have these discussions every trip we took probably where [the implementer] would want to go into the late groups and we would just tell them they couldn't...It was incredibly tedious work...the number of times we had to have a conversation about them not entering the other areas kind of shocked me" (Interviewee 3E, personal communication, June 24, 2016).

In Ghana, this type of interaction was continuous because of the evaluator's local presence. ISSER's local presence allowed both early engagement and trusted communication. As one expert who helped conduct the evaluation said, "I think that it is the early engagement with those on the ground that may have contributed to the low contamination" (Interviewee 3O, personal communication, August 2, 2016). According to someone at MCC involved with the evaluation, "I think threats to the control group was really managed by ISSER. We just had this great luxury. It's very rare that you had a kind of ISSER entity on the ground that had such a good relationship with the Ministry of Agriculture and all the other implementers. There was a level of trust there" (Interviewee 3H, personal communication, June 29, 2016).

ISSER's early engagement and local presence successfully maintained the control group for the randomized rollout design. As a result, the temporal and spatial sequencing of the training was contrary to how some MiDA agricultural specialists, implementation consultants, and local leaders and farmers would have liked to see it evolve. According to one agricultural specialist, "Within the MCA we had to make accommodations, so many accommodations for the concerns about the contamination...sequencing actually was probably the biggest effect of the evaluation. Sequencing of implementation had to go hand in hand with the evaluation and that complicated the implementation. It stretched it out" (Interviewee 3J, personal communication, July 6, 2016). In addition to the timing of implementation, the randomized rollout also forced implementers to

spatially implement training in an inefficient way: “In a geographic way the selection was random, it meant that you had to go to different places at different times so you couldn’t plan sequencing, if you like, in a straight line” (Interviewee 3O, personal communication, August 2, 2016). Both the randomized rollout’s temporal and spatial consequences on Ghana’s FTT led for inefficient implementation that also likely raised project costs: “To the extent that the randomization meant that you couldn’t geographically plan the sequence of the program, it meant that it was a bit more costly for the program” (Interviewee 3O, personal communication, August 2, 2016).

The Ghana FTT was part of an integrated, \$200 million agricultural and rural development program that included infrastructure development through irrigation and roads and value chain development through post-harvest centers and credit. The value chain approach to agricultural development is an integrated, systems-based approach. By conducting a randomized rollout evaluation and influencing the implementation of just one piece of this integrated approach, there was a risk of affecting the implementation and efficacy of the other approaches (Interviewee 3J, personal communication, July 6, 2016). The randomized rollout evaluation did not affect the irrigation or roads projects, but it did affect the credit project. Credit flowed to farmers who had developed business plans, and business plan completion was an outcome of the FTT. One of the evaluators conceded that “[f]or the credit component, which had to use farmers that were being trained because of the business development plans etc. that were part of the training, [credit] may have followed the randomization” (Interviewee 3O, personal communication, August 2, 2016).

ISSER’s early engagement and close relationship with MiDA and “intensive and frequent” communication with implementers ensured little to no contamination of the control

group. There was one exception to this, however. In the Southern Horticultural Belt, ISSER detected some contamination:

There seem to have been some level of contamination of the control group. This was a problem that the farmers in the southern zone particularly raised strongly. There were two sources of this contamination. One was from the control farmers attending training sessions meant for the treatment group. The other source was engendered by the situation where farmers who got the training went around to their colleagues in the control group...and taught them what they had learnt (ISSER, 2012: 33).

ISSER estimates that five percent of the late treatment group was trained as part of the early group. While this contamination was neither large nor fatal to the statistical design, it does shed light on the limits of social control around RCTs.

Few evaluation economists or those actors they enrolled and mobilized across the network took into account project participants' agency. But ignoring participants' agency was not a luxury implementers could afford. Field technicians working for implementers or extension agents working with Ministries of Agriculture that worked with implementers had longstanding relationships with farmers from previous projects. They were also often farmers' neighbors. They bore the brunt of farmers' grievances, and in turn, voiced these concerns to their supervisors.

Farmers' interests were based neither in accountability nor learning for development impact. Their interests were material. Farmers are usually some of the lowest income citizens in countries in the Global South. They wanted training and inputs as soon as possible. As a Ghanaian saying goes, "If the rain begins to fall, the earlier you put your bucket to catch the rain, the better" (Interviewee 3N, personal communication, August 1, 2016). Just as one does not know when the rain will stop, one cannot predict when project funding will cease: "I mean, everybody wants it today because you cannot be sure that tomorrow it will happen. There have been instances of programs that have started and somewhere down the line a donor or the

government has said ‘I’ve run out of resources’ ” (Interviewee 3O, personal communication, August 2, 2016). A big motivation for farmers to clandestinely attend early trainings was provision of the starter pack with seeds and other material inputs, which was valued at over US \$200. In Ghana, there are substantial regional differences in rural wealth and some of the farmers were so food insecure that seeds from starter packs were consumed immediately rather than planted (Interviewee 3I, personal communication, June 30, 2016) (seed consumption is a survival strategy during periods of food insecurity).

Contamination was higher in the Southern Horticultural Belt because there were stronger network ties and forms of social capital among farmers (Interviewee 3M, personal communication, July 27, 2016). Farmers in the early group would communicate about training activities to late treatment FBO members and some would sneak into early treatment trainings. Strong social capital and networks enhance the public goods nature of private, individualized interventions such as training and technical assistance. Consequently, they may improve project outcomes but undermine a randomized evaluation design.

### **The End Life of RCTs: Indeterminacy and Lack of Impact**

On July 20, 2012 MCC held an *Agriculture Evaluations Portfolio Peer Review Workshop* to “critically assess the evaluation results of five MCC-financed agriculture programs...through peer review” (MCC, 2012a: 1). These completed FTT randomized rollouts were dubbed “The First Five” for having been MCC’s first completed RCTs. They were subjected to “peer review” by five academic and professional economists such as Yale University economics professor and J-PAL affiliate Chris Udry and World Bank economist Markus Goldstein. The peer reviewers submitted their assessments and, together with the evaluations, they were published online. On October 23, 2012 MCC released a report summarizing the conclusions of the evaluations and

peer reviews called *MCC's First Impact Evaluations: Farmer Training Activities in Five Countries* (“Brief”) (MCC, 2012b).

The findings around the FTT project outcomes were either inconclusive or negative. Where RCTs did not hold together for the course of the project, as in the Honduras FTT, results were indeterminate. In these evaluations, the methodology was compromised (i.e., the control group was invalidated) and MCC could not trust the estimated impacts. Where RCTs were successfully completed, like in Armenia, Ghana, and Nicaragua, the evaluations were unable to detect changes in household income. As one MCC official said, “So here was our chance, we were getting impact evaluations that were gonna be able to show the world how much we had raised incomes; none of them were able to. So then it was like, crap...the whole world that cares anything about MCC is watching and we’re gonna have to show that we didn’t achieve what we wanted to achieve or we don’t know whether we did because the evaluations don’t tell us that (Interviewee 3C, personal communication, May 26, 2016). Some development economists critical of RCTs would lay blame on the method’s rigidity. Dani Rodrik called RCTs “straightjackets” that cannot be adapted once set in motion (Rodrik, 2008). MCC evaluation economists’ decision to focus on income without using a research design that allowed enough time to observe changes in household income could be blamed for lack of impact. Save for the Nicaragua FTT, there were limited effects on outcomes such as farm incomes and adoption of improved techniques as well along the program logic chain.

Indeterminacy can be troubling to those who have been physically and emotionally involved in a project: “The [Ag team] was like ‘oh my god, this is shaking us to our knees’ because things that we thought were working are not working or we don’t know whether they’re working because we didn’t design the damn thing right or we’ve got a complete mismatch”

(Interviewee 3C, personal communication, May 26, 2016). Lack of perceived impact can be downright devastating. An MCC senior development policy official described the reaction: “I mean, we did the peer review conversations for those impact evaluations and the Ag team was in the room and [the Sector Director] was in tears because s/he had put blood, sweat, and tears for six years designing and overseeing the implementation of the agriculture project in [country] and six years later s/he had no [expletive] idea whether it had achieved anything” (Interviewee 3C, personal communication, May 26, 2016).

Ultimately, it is impossible to tell if FTTs achieved their stated aims. This is because RCTs don’t just observe; they intervene. Evaluations lead to what I call a disposition for a Type I error, or failure to reject the null hypothesis that the project had no effect, because when successfully implemented RCTs can inhibit project outcomes from materializing. They intervene in various ways while also appearing to capture the resulting lower impacts in a supposedly rigorous and credible way. As a result, RCTs make it appear as if the FTT, or any project intervention, was not effective when in fact it was the combination of the project *and* the RCT that contributed to project outcomes. If an RCT is responsible for participant selection, interfering in projects’ operational efficiency, and affecting related components of integrated programs, then they cause, rather than neutrally observe at a distance, project outcomes. One of MCC’s evaluators framed the tradeoff RCTs presented as thus: “We had to destroy the program in order to evaluate it” (Leonard, 2011).

The Brief on the First Five FTT evaluations erased any doubt that an RCT was more akin to a project than a study. What was remarkable about this report was how little farmer-training projects were discussed. It instead spent as much time talking about lessons learned around RCTs’ implementation (MCC, 2012b). The conclusions reached attested to the fact that they

were not evaluations but interventions. Like the development projects they aimed to evaluate, RCTs too faced implementation challenges that required political work such as enrolling allies and stabilizing networks thought to be the reserve of non-scientists to overcome.

Neither indeterminacy nor lack of impact among randomized evaluations is new. A review of eight community level nutrition interventions implemented in the 1970s concluded, “In no case could we make an unambiguous, internally valid attribution of change in nutritional status to the project’s intervention” (Schön, Drake, & Miller, 1984: 12). According to the authors, indeterminacy is the result of implementation contexts tending to change without regard for the experimental design (Schön et al., 1984). When results are determinate, they tend to show no program impact. This was particularly true of RCTs. This is why Schon and Rein (1995) call such assessments “program killers”.

The results of the First Five therefore put evaluation economists and their agricultural specialist allies in a bind. With accountability minded managers and external audiences such as Congress awaiting the results, how would MCC manage to explain that \$180 million dollars worth of investments had no impact? Or more accurately, that approximately \$10 million dollars spent on randomized evaluations could not conclusively say if any impact was achieved?<sup>49</sup>

## **Conclusion**

This chapter has shown how MCC evaluation economists transformed the academic purposes of the RCT for development evaluation and the MCC results agenda. Senior officials welcomed RCTs for their ability to demonstrate the impact of MCC investments to an external audience. The stabilization of RCTs was greatly aided through CGD’s network of economists

---

<sup>49</sup> \$10 million was sufficient to train approximately 8,000 farmers. In other words, for the cost of the five RCTs, MCC could have funded an additional Nicaragua FTT.

and evaluation experts. Through its working group, the CGD established a hierarchy of evidence in project evaluation. This made it more difficult for performance evaluators to argue the merits of their preferred input-output approach to evaluation that lacked a counterfactual. Other opponents, such as sector specialists, had to be dealt with more directly. Evaluation economists did political work to cast RCTs as tools for learning rather than accountability, which appealed to some sector specialists' duty towards achieving best practice in their sector and improved outcomes for project participants. Sector specialists' and other important stakeholders' opposition was further softened in evaluation workshops.

Having sector specialists consent to RCTs was essential to overcoming opposition from implementers. In farmer technical assistance and training projects—where most RCTs were conducted because of the method's bias towards evaluating projects delivering private, individual goods—implementers wished to target older, better off, and more educated farmers first because they were pioneers and would facilitate the diffusion of innovation and learning. Where RCTs prevailed, farmers had to be targeted randomly for early or late treatment. This had the political effect of leading the “average farmer” in both groups to look similar. Randomization also intervened in the implementation process by forcing the implementer to maintain the control group. This was an act of control, or power, that changed how the implementer sequenced training temporally and spatially. RCTs' intervention in project participant selection and implementation led to either indeterminate or suppressed project impacts. In other words, the RCT ended up damaging the project in order to evaluate it.



## Chapter 7: The Power of State Expertise

### Introduction

This chapter will first summarize the three cases of the dissertation and then provide an interpretation of what the empirical findings mean for how economic methods are stabilized in the state and the effects they produce. I use the term *stabilized* because institutionalized gives a false sense of permanency. Economic methods as governing rationales are contingent, contested, and open to future (re)consideration. The cases show that, if we “open-up” the state, we witness conflicts between different groups of state experts and between their rationales. State economists faced alternative rationalities as they attempted to stabilize their governance methods. The dissertation contributes to theories about the power of state expertise by combining empirical evidence of how economic experts achieved autonomy and overcame opposition from other expert groups from the cases with modern and postmodern theories of power. An economic expert’s actions are motivated as much by political interests as any other actor’s are. State economists have an interest in stabilizing their methods as governing rationales—in other words, rationality is politics by other means. The methods are central to economists’ political work and the level of power they require to fulfill their political work is commensurate to the level of opposition they face. Governance methods have powers independent of the actors that promoted or authorized their use. As forms of power, methods have political effects that are not simply discursive. These effects include the reshaping of bureaucratic power relations, a shift in decision-making processes and actors, and the distribution of global development finance resources.

## Summary of Dissertation Cases

The first case study examined how the Millennium Challenge Account (MCA) employed governance indicators to decide which countries to award large development finance grants. New institutional economics introduced a new explanatory variable to the neoclassical economics toolkit—governance. Neoclassical macroeconomists at international financial institutions in Washington, DC black-boxed the new ideas of institutional economics by equating them with proxy indicators of governance derived from private-investment risk-analysis firms. By adding these governance indicators as a new explanatory variable for growth to their cross-country growth regressions, neoclassical economists were able to speak with greater authority about a topic normally reserved for other experts. Neoclassical economists then brought together governance indices with research on aid effectiveness to analyze the influence of governance on economic growth. This was a major catalyst for the development-aid strategy of selectivity—choosing countries based on their performance according to a set of defined quantitative indicators.

Selectivity through governance and other indicators helped the Bush administration bridge a widening gap between global- and civil-society demands for increased development finance and rising suspicion of development aid among conservative lawmakers. The use of governance indicators in econometric research by World Bank economists played an important role in stabilizing selectivity. Economist Steven Radelet, a US Treasury official and Center for Global Development (CGD) Senior Fellow, was a key figure in a network that spanned the World Bank, CGD, and federal government and defined many of the parameters of the MCA selection system. In particular, his relationship with the creators of the World Bank's Worldwide

Governance Indicators (WGI), Daniel Kaufmann and Aart Kraay, was an important factor in the decision by the MCA team to consider the WGI for selectivity.

Use of the WGI, in addition to other indicators to distribute aid funding, was a novel, (neo)liberal internationalist approach to bilateral aid. This approach was at odds with the conflicting rationality of political realists who distributed foreign aid through a geopolitical rationale. The economic policy community advocating for the MCA had to overcome the foreign policy community's resistance and did so by developing a strategy of "trust in numbers". Through this strategy its members cast the foreign policy community as political and subjective while casting themselves as apolitical and objective. The economic policy community enrolled conservative legislators and White House Office of Management and Budget (OMB) officials who shared a commitment to the selectivity strategy and curtailed the discretion of the foreign policy community.

Despite the economic policy community's successful "trial of strength" (Latour, 1988) against the foreign policy community, differences over the specific indicators used in the selection system remained. In particular, academic economists in the economic policy community were skeptical of the WGI's aggregation methodology and thus its measurement validity as well. Treasury Undersecretary John Taylor proposed an alternative to the WGI that Treasury economists had created, but this index was rejected because it was not "transparent" in the way the WGI was. The CGD had defined transparency as a selection criterion being public while research macroeconomists believed transparency was rooted in measurement validity. CGD's criteria, along with the WGI's broader country coverage, meant that the WGI was more suitable for operationalizing the MCA's global indicator-based competition for funds. When it

came to the WGI's use in the selection system, policy expediency trumped methodological concerns.

The WGI's use had an effect on the nature of bias in decision-making as well as the distribution of development finance. Decision-making was shifted away from the Washington, DC foreign policy community and towards an assemblage of country risk analysts, survey instruments, and statistical models. In the end, the WGI's incorporation as a decision-making tool for aid allocation did not lead to rationality's taming of subjectivity but swapped the subjectivity of foreign policy experts for the subjectivity of a diverse and distributed network of mostly country investment risk analysts. It collapsed incredibly complex country politics into a singular metric that ignored important factors for governance such as years since independence and institutional path dependencies. The indicators included another type of bias—income bias. Wealthier countries routinely performed better on the WGI than did relatively poorer ones. Countries that passed four or more WGI indicators and qualified for MCA assistance were on average, wealthier than those that did not. Because WGI scores barely changed over time, this resulted in a set of countries that became a “global underclass” when it came to eligibility for global development finance.

The second case examined the Millennium Challenge Corporation's (MCC) use of growth diagnostics to determine in what kinds of sectors and projects the development agency should invest. In the wake of dissatisfaction with structural adjustment's prescriptive programs, three Harvard economists devised a decision-tree framework that assessed constraints to private investment using four data-driven empirical tests based on a Bayesian process of elimination. The growth diagnostic was meant to tailor reform programs to country-specific contexts while adhering to the principles of neoclassical economics. MCC designed nearly twenty bilateral

investment agreements, or compacts, before it began using the growth diagnostic. Early compacts were developed in an ad hoc way between MCC senior officials, sector specialists, and a partner government's "core team" comprised of officials from the prime minister's office and various ministries. Country proposals for compact funding were often informed by or directly based on Poverty Reduction Strategy Papers (PRSP), a vehicle for donor assistance governed jointly by the World Bank, International Monetary Fund, and the country government, using varying degrees of civil society input.

MCC economists were uneasy with this supposed informal approach to compact development that produced "Christmas tree" compacts with many disjointed parts from country proposals full of pork barrel spending. Economists also felt that MCC-sector officials were not at all objective in their negotiations with the core team and steered projects towards their own sectors. MCC economists pitched growth diagnostics, or constraints analysis, to bring an "objective" framework to this "politicized" process. Senior officials meanwhile looked for a solution to what was a laborious compact development effort—they wanted a way to quickly winnow proposals into compacts.

Growth diagnostics are congruent with the Finance Ministry tendency held by the finance ministry group (comprised of members of finance ministries, economic policy managers, economic analysts, and orthodox academic economists). At MCC, the economists were closely aligned with the Finance Ministry tendency. Growth diagnostics contended with the Civil Society tendency held by the civil society group (comprised of a broad range of advocacy and NGO groups, members of certain United Nations specialized agencies, social sector ministries, and many academic non-economists). At MCC, the Environmental and Social Assessment team (ESA) was closely aligned with the civil society tendency. The main points of contention

between the two groups were over whether and how to address inequality. The faith of MCC economists in competitive markets as a solution to growth made them more tolerant of inequality. They believed that competitive markets and market-led development could resolve structural inequalities over the medium to long term. The civil society group believed that structural inequalities affecting social institutions also affect markets: markets cannot resolve inequalities on their own, and marginalized groups need to be purposefully targeted. National-level sectoral reform programs tailored to aggregate private investment would not suffice.

Despite their conflict with civil society rationales, constraints analyses were incorporated into compact design with little to no resistance. This was because constraints analyses received senior officials' blessing and the ESA team's work was mostly about reducing harm (e.g., environmental permitting, proper resettlement) rather than proactively addressing inequality. Even if the ESA team took a proactive role, its members were scattered across the agency, which made it more difficult to build a network at MCC and with like-minded social ministries in partner countries to resist constraints analyses by economists. MCC's adoption of constraints analysis elevated economists in the compact development process. Sector specialists were demoted and relegated to the role held earlier by economists—data providers.

By 2015, MCC constraints analyses informed fourteen compacts. MCC economists led their analysis teams that included anywhere from a handful to a dozen economists, analysts, and consultants from Washington, DC, partner country governments, donors and development banks, and the Harvard University center where growth diagnostics originated. This team identified, on average, three binding or highest priority constraints to growth. While MCC economists decried the subjective biases of sector specialists, bias had now shifted to the decisions that economists made regarding diagnostic analysis such as which constraints to consider and test. Constraints

regarding infrastructure, micro risks to private investment, and human capital were tested most often. Data availability and quality were often themselves a constraint on what could be reasonably examined. Women, the informal economy, and rural areas were often invisible to the constraints analysis team because of data paucity and the existing beliefs of analysts.

Constraints analyses had important impacts on MCC's operations and investments. The constraints analysis substituted for the development finance community's Poverty Reduction Strategy Paper (PRSP) and, on average, reduced the length, breadth, and depth of participatory processes. PRSP consultations typically gave voice to the rural poor. Constraints analyses were, by contrast, dominated by economists and dismissive of rural concerns. As intended, they concentrated investments in a smaller number of sectors and projects. Because of their neoclassical economic approach, constraints analyses shifted investments from rural to urban areas. Their focus on reform shifted investments from "hard" infrastructure to policy and institutional reform. MCC compacts began to resemble the World Bank's sectoral structural adjustment programs of the past. Finally, they shifted discretionary authority away from presidents and ministers in partner countries to White House and MCC officials in Washington, DC. Energy, rather than Transport or Agriculture and Rural Development, became the dominant sector in MCC's investment portfolio in large part because of the Obama administration's Power Africa program, sometimes in contrast to the priorities expressed by country officials.

The third case examined MCC's use of randomized controlled trials (RCT) to conduct impact evaluations of development projects. MIT's Poverty Action Lab (J-PAL) made RCTs popular by the early 2000s after economics professors Esther Duflo and Abhijit Banerjee partnered with small non-governmental organizations in India to experiment with education projects. Around the same time, CGD set up an evaluation working group to review how the

development community conducted project evaluation. The working group, of which Duflo was a member, established standards of evidence. Based on these standards, it determined that there was a significant “gap” between using counterfactuals to assess impact—of which RCTs represented the “gold standard” approach—and evaluations that development agencies were conducting.

A network formed around RCTs consisting of J-PAL, the CGD evaluation working group, and the Bush administration’s interagency group that created the MCA. A key member of this interagency group, Delia Welsh, became part of the network during her tenure at MCC and introduced RCTs to MCC’s monitoring and evaluation (M&E) department as a way to implement MCC’s results agenda. With the help of CGD’s establishment of evidence standards, MCC evaluation specialists transformed the academic purposes of the RCT for development evaluation. RCTs were incorporated into the M&E department despite resistance from “traditional” performance evaluators because MCC senior officials found them appealing for reasons of accountability.

RCTs faced other sites of resistance in addition to the M&E department. Sector specialists were reluctant to adopt RCTs because they interfered with project implementation and also held them accountable for things outside of their control. MCC evaluation economists cast impact evaluations as serving either senior officials’ goals of accountability or sector specialists’ goals of learning depending on which group they were working with. To win sector specialists and other implementation team members over, RCT advocates held workshops where they promoted, taught, and preliminarily designed RCT evaluations. By 2015, MCC had initiated 23 RCTs, accounting for one out of every six impact evaluations. RCTs were not, however, distributed evenly among MCC’s investments. MCC impact evaluators sought out projects that



were most amenable to RCTs: training programs delivered to individuals. The structure of the RCT makes it nearly impossible to apply them to public goods and networked infrastructure like roads and easier to randomize projects with individually delivered benefits such as training to poor farmers. The most common MCC sector with RCTs was Agriculture and Rural Development.

RCTs were presented as research studies but in reality they were parallel projects that unfolded alongside a development intervention throughout implementation. They randomized project participants into treatment and control groups, and these groups needed to be maintained over the project's lifetime. In farmer training projects, choosing, or targeting participants randomly conflicted with purposeful targeting based on the project's program logic. This logic was geared toward maximizing crop productivity and income, necessitating project participants who were most likely to succeed at farming. The targeting criteria of project implementers were based on the diffusions of innovations paradigm and included a set of observable (e.g., assets, education level) and unobservable (e.g., motivation, skill) factors. The project implementers' use of unobservable farmer characteristics for project participation was particularly problematic for impact evaluation because it confounded implementer judgment with the training itself. To attribute impact to the project intervention—a key objective of accountability—MCC impact evaluators had to ensure that project participants were selected randomly.

RCTs successfully randomized project participants in four of the five farmer training projects examined. In these cases, the evaluation team obtained the consent of agriculture and rural development sector specialists, MCC country directors, and country government officials to implement an RCT. Contracts between MCC and project implementers mandated that implementers cooperate with impact evaluators' RCT designs. When participants were

randomized, treatment and control groups closely resembled each other on demographic and economic characteristics. Implementers, however, would have wanted the treatment, or early training group, to be better educated and wealthier.

Randomizing participants was only the first step; the control group required maintenance over the project's lifetime. This meant limiting contamination, or contact, between treatment and control group participants. Units such as farmers, farmer organizations, or farming communities were sampled to maintain spatial and temporal distance. This also presented tensions with the implementers' rationale: implementers wanted to encourage contact between "pioneers" and other farmers to boost project outcomes through improved innovation diffusion and train everyone as soon as possible rather than waiting to train the control group. To keep implementers from training control group farmers sooner rather than later, the evaluation team had to continuously negotiate with and persuade the implementers.

When the control group was successfully maintained, it affected project implementation. Implementers altered how they would normally sequence their training temporally and spatially. In some cases, costs of implementation rose to accommodate the RCT design. In integrated programs where farmers received financial credit following the completion of training, the RCT could also affect when and to whom loans were distributed first. Even when evaluation teams contractually obligated implementers to abide by random selection and successfully persuaded them to maintain the control group, instances of contamination, albeit limited, still occurred because project participants' concerns were neither accountability nor learning (from impact evaluations). Instead, poor farmers were motivated to acquire resources.

The two most common results of the five farmer-training RCTs were indeterminacy or lack of impact. These are indicative of the fate of RCTs at MCC more generally. Where

evaluation teams and the implementer could not maintain the control group, the RCT evaluation's findings were judged indeterminate. Where these efforts were successful, RCTs were "successful" and able to document project impact. Those impacts were lower than they might have been without the RCT because it interfered with implementers' rationales for participant selection and implementation sequencing. The RCT ended up altering the project's original strategy in order to evaluate it.

### **The Power of State Expertise**

Let us briefly revisit the issues and questions motivating this study. A number of scholars, particularly social reformers and development critics on the left and right, have failed to describe certain forms of state rationality such as neoliberalism or managerialism as political accomplishments because they took the power of intellectual rationality as predetermined. They also give disproportionate attention to experts' discourse and ideas while neglecting their methods. This dissertation posed questions to address these shortcomings. First, I asked, how do certain economic methods, or policy devices, get adopted by state organizations as governing rationales? Second, I asked, how do these methods affect power relations, decision-making, and the distribution of resources once stabilized?

My answer is that the stabilization of economic methods is the result of political work. This involved economists exercising various strategies of power to gain autonomy (see also Dargent, 2011). In doing so, experts faced resistance from other expert groups with conflicting rationalities. This happened at different "sites" across the cases (see Table 7.1). The main site of conflict lay within state policymaking and administrative bodies such as the National Security Council (NSC) and MCC. The other major site of conflict was the project implementation context, which was relevant to the case of RCTs. In the case of governance indicators, the

economic policy community and its neoclassical economic rationality came across opposition from the foreign policy community’s geopolitical realists in the NSC interagency working group. In the case of growth diagnostics, while the civil society community was concerned about the implications of MCC economists’ growth rationale, it was not in a position to present opposition to the growth diagnostics of economists when they were first introduced. In the case of RCTs, evaluation economists came across opposition at two sites—in the Compact Transaction Team and project implementation environment—from sector specialists and implementers who had different rationales for the selection of program participants.

Table 7.1

*Sites of Conflicting Rationalities*

	Sites of Conflict	Main Sites of Conflict	Economic Rationality	Alternative Rationality
Growth Diagnostics	0	None	Ministry of Finance Tendency	Civil Society Tendency
Governance Indicators	1	Interagency Working Group	(Neo) liberal Internationalism	Geopolitical Realism
RCTs	2	Transaction Team Project Implementation	Experimentation via Randomization	Diffusion of Innovations

As a result, economic experts had to do political work—the kind often thought to be the exclusive reserve of politicians and similar actors with vested “interests”—to overcome opposition to their methods from other state experts with alternative rationalities.

After becoming stabilized, methods possessed not only the “small-*p*” political power to define, quantify, standardize, surveil and normalize, as many sociologists and governmentality

studies scholars have alluded to, but also “big-*P*” political power to transform organizational power relations and determine the distribution of resources. The remainder of this chapter outlines the political work of experts and these small-*p* and big-*P* political effects and concludes with thoughts on how theories of power might be employed to study state experts.

### **Gaining Authority by Achieving Autonomy**

In the documented cases, both state economists and senior officials were occupied with autonomy. State economists wanted to stabilize their methods and senior officials sought to boost the bureaucratic autonomy of their respective departments to shield them from external public and legislative scrutiny (Carpenter, 2001; Huber & Shipan, 2002). Senior officials welcomed any strategy that could improve their department’s legitimacy (DiMaggio & Powell, 1983). In response, state economists enrolled senior officials in their entrepreneurial efforts by arranging solutions to organizational problems, mediating between the worlds of knowledge production and application, and framing themselves as the ideal spokespersons for objectivity.

**Arranging Solutions.** Social scientists studying state expertise who focus on expert autonomy generally suggest that knowledge alone is insufficient for expert power; power also requires autonomy (Centeno & Silva, 2016). In all my cases, state economists relied on the consent of and autonomy from superiors in order to introduce and stabilize their method. NSC officials allowed economists in the US Treasury (“Treasury economists”) to devise a country selection system that included governance indicators. Officials of MCC Compact Development permitted compact development economists (“MCC economists”) to execute growth diagnostics for compact design. Department of Accountability officials at MCC agreed to let evaluation specialists (“evaluation economists”) implement RCTs as part of MCC’s results agenda. Unlike

existing studies of expert autonomy, methods were central to experts' ability to achieve autonomy.

In order to gain autonomy, state economists had to enroll their superiors in their efforts to stabilize methods for decision-making. Senior officials identify broad policy objectives but rarely have the foresight, expertise, or time to identify how their objectives will be realized (Grindle, 1977; Justesen & Mouritsen, 2011). These are opportunities for experts to gain autonomy. Broad policy objectives such as "aid effectiveness" in the case of governance indicators or "results" in the case of RCTs are usually formulated in response to external pressures demanding something of senior officials or their organizations. Except in rare cases, senior officials are unable to deliver on these demands without the technical know-how of expert colleagues (Grindle, 1977). Thus, expert bureaucrats are in a position to translate the programmatic ideals of policy into administrative and governance practice (Hart, 1992; Huber & Shipan, 2002). State economists translated the interests of these superiors by being what Benveniste (1972) calls "arrangers of solutions" to the problems that senior officials faced.

In each case, some of the important entrepreneurial work that state economists accomplished to gain autonomy involved creating a "fit" between the intellectual purpose of a development-economic method and the development governance problem to which that method could be applied. Senior NSC officials, for example, consented to a strategy of selectivity but were content to leave the details up to the interagency group. Gary Edson, the Bush administration's Deputy National Security Advisor for International Economic Affairs, knew that he wanted a way to select countries based on their policies and institutions but was not fluent in the details of specific indicators. This gave Steven Radelet and his team considerable leeway in determining the details of the selection system. Senior officials in MCC's Compact

Development department wanted a way to streamline country proposals and the compact development process, but they did not present their own solutions for doing so. The proposal to use growth diagnostics came from MCC economist Ben Dennis following his consultation with his economist colleagues. Senior officials in MCC's Department of Accountability needed to respond to external pressures to measure and demonstrate results, but they were somewhat at a loss to stipulate what that entailed. MCC's evaluation economists partially filled the vacuum with a proposal to use RCTs for project evaluations.

State economists were able to propose solutions because they had knowledge of development economics methods appropriate to their field. Treasury economist Steven Radelet, MCC economist Ben Dennis, and evaluation economist Delia Welsh belonged to epistemic communities, or knowledge-based networks. In the case of governance indicators, the core nodes of this network included the Treasury and the World Bank; in the case of growth diagnostics, the Harvard Kennedy School of Government and MCC; and in the case of RCTs, J-PAL, the Center for Global Development, and MCC. State economists like Radelet, Dennis, and Welsh became familiar with economic methods like the WGI, Growth Diagnostics, and RCTs in their areas of expertise (macroeconomics, development economics, and program evaluation, respectively) through contacts with research economists in these intellectual centers. This is precisely how some sociologists have suggested expert methods get introduced into state organizations: “[e]conomic expertise is not something that is formulated in the academy and then ‘applied’ elsewhere. It is a network that stretches from the academy through businesses and quasi-governmental organizations, along a permanent port into the state” (Eyal & Levy, 2013: 232). As was the case in economic policymaking in Colombia and Peru (Dargent, 2011), this scientific

authority helped to enroll senior officials and boosted the autonomy of experts by presenting a solution whose details the politicians could not immediately or easily grasp.

The combination of state economists' organizational efforts to stabilize their methods and ability to draw on research methods in development economics merits a new category of institutional entrepreneur: "methodological entrepreneurs." When discussing the introduction of new policies in an institutional field, institutional theorists call these pioneers "policy entrepreneurs" (Mintrom, 1997). Methodological entrepreneurs represent a specific type of expert entrepreneur, one who aims to change the way institutions do work based on their particular interest in stabilizing methods for governance and decision-making.

**Mediating to Manage Dissension.** Proposing and arranging solutions for senior officials' organizational challenges was only the first step in gaining autonomy. There were competing and compelling alternatives from other credible economic experts for senior officials to consider. State economists managed dissension to their proposals from disciplinary peers by mediating between the worlds of knowledge production and its application (Stehr & Grundmann, 2011). Mediation involves the use of methods as instruments rather than principles. Those who mediate are motivated by "syntheses that work" among important constituencies such as senior officials, research economists, and economists "in the wild" and are as expert in the contexts in which methods are deployed (such as development finance agencies) as the methods themselves (Osborne, 2004). In the case of governance indicators and RCTs, state economists successfully mediated between two attitudes towards statistics—"metrological realism," or a commitment to methodological validity, and "proof-in-use," or a commitment to policy expediency (Desrosieres, 2001). Those concerned with methodological validity are committed to reliability in terms of accuracy and precision; they adhere to the "moral economy" of science that involves



quantification, empiricism, and objectivity (Daston, 1995). Those concerned with policy expediency display little interest in the origins or processes that generate methodological outputs (Desrosieres, 2001).

In the case of governance indicators, the head of the indicator working group, John Taylor, was preoccupied with methodological validity. He warned NSC senior officials about using the World Bank's Worldwide Governance Indicator (WGI) database to rank countries, objecting to its data aggregation methods and high standard errors. Meanwhile, NSC senior officials needed a set of governance indicators that could measure and rank as many countries as possible. These officials were more interested in the WGI's ability to yield a specific point estimate for a large number of countries for the purpose of comparison than they were in any debate over the validity of the WGI. Treasury economists mediated between Taylor's commitment to measurement validity and NSC officials' desire for policy expediency. They advocated for the WGI's adoption because they felt that a commitment to metrological realism was unsuited to the selectivity strategy. Ultimately, dissenters like Taylor had to accept NSC officials' attitude of policy expediency and the use of WGI for governance purposes it was neither designed nor necessarily well suited for.

In the case of RCTs, evaluation economists appreciated the need of senior officials to "focus on results" but were also keenly aware of the methodological concerns J-PAL economists had about existing evaluation practices. Performance evaluators, however, were skeptical of the emphasis that J-PAL economists placed on the establishment of a counterfactual via randomization. They believed that evaluating project outputs against the level and quality of organizational inputs could provide an adequate project performance assessment. MCC senior officials had little interest in or concern about these debates. Their attitude towards evaluation

was one of policy expediency—whichever evaluation approach would best portray MCC’s commitment to results was most appropriate. In this case, evaluation economists felt that the policy expediency of performance evaluations did not suit MCC’s results agenda. Their mediation resulted in the belief among MCC senior officials that randomization would deliver attribution of impact to MCC investments. Senior officials believed that such definitive rigor would make MCC a pioneer in the field and boost the agency’s standing as a leader for accountability with legislators and peer organizations in the field of global development finance. Performance evaluation dissenters ultimately had to accept the methodological standards set by J-PAL economists as a core feature of MCC’s evaluation strategy.

State economists’ participation in think tank–led expert networks was key to managing peer dissension. Their proposed methodological solutions were simultaneously gaining establishment as field level standards in global development finance through CGD networks. As Brunsson and Jacobsson (2002) state, “Standards may productively substitute for various other forms of authoritative rule...although standards are often promulgated by experts, they may come to function as... a way of embedding authority in rules and systems” (Brunsson & Jacobsson, 2002: 32). This external source of legitimate authority made the proposed methods “safe” choices for senior officials.

The Center for Global Development assembled a network of specialists from the US government, non-governmental organizations, development research communities, and international financial institutions across five workshops in Monterrey, Mexico, and Washington, DC to, in part, establish “optimal characteristics of data used for [MCA] eligibility criteria” (Birdsall, Levine, et al., 2002). The think tank also assembled a network of academics, evaluation and policy specialists, and state economists into an evaluation working group that

established standards of evidence where evaluations employing counterfactuals were superior, especially those established through randomization, to those that did not (The Evaluation Gap Working Group, 2006).

State economists' participation in CGD networks ensured that senior officials would not perceive economists' proposed methodological solutions as idiosyncratic projects tied to individual professional or personal interest but rather as projects adhering to emerging field-level standards. The adoption of these methods *cum* standards by risk-averse senior managers bolstered MCC's legitimacy and standing among legislators, civil society institutions, and other development finance organizations. This improved the chances of senior officials to gain bureaucratic autonomy for their respective departments. State economists' mediation and participation in expert networks drew senior officials down the path of the WGI and RCTs while cutting them off from alternatives such as Taylor's proposed alternative governance index and performance evaluations.

**Framing and Shaming.** The final strategy state economists employed to gain autonomy was presenting themselves as the ideal spokespersons for objectivity and thus free of personal or political bias. This is an important strategy of power economists employ (Hirschman & Berman, 2014; Porter, 1996). State economists cast themselves and their methods as more objective than existing alternatives. This was an act of framing that involved generating tensions around the validity of existing decision-making arrangements and matching new projects to the interests, values, and problems of potential allies (Garud et al., 2007). As in other cases of framing in organizations, the goal was to convince potential allies such as hierarchical superiors that alternatives were valid, appropriate, and indispensable (Rao, 1998). State economists questioned

the legitimacy of existing decision-making processes and attempted to persuade senior officials that their more objective approaches were in everyone's best interest.

In all the cases, economists framed existing decision-making as rooted in participants' personal self-interest and those involved as having "selection bias" in their decisions on which countries, sectors, and project participants to fund. Foreign policy officials had their "favorite" countries. Country governments and sector officials had their "pet projects". Implementers and local officials "cherry picked" project participants. In each case, these actors were purportedly partisan and subjective while the economic approach was rational and objective.

Existing studies suggest that economists can effectively claim to be spokespersons for objectivity because of their quantitative methods (Hirschman & Berman, 2014; Porter, 1996). As standards, methods have the character of procedural objectivity and as standardized processes are more transparent in ways that support accountability (Fuchs, 1997; Timmermans & Epstein, 2010). However, a potent and seemingly overlooked source of economists' claims to objectivity is that their discipline is the most nomothetic among the social sciences (Wallerstein, 1996).<sup>50</sup> State economists were able to cast themselves as objective in large part because they presented their interests as rooted in universal laws that transcended the self rather than in particular self-interest or partisan relationships (i.e., politics). In other words, they argued that they were the only experts with aperspectival objectivity (Daston, 1992). In the cases studied, state economists believed that universal laws, sometimes stated in the form of stylized facts—for example, better governance leads to greater national growth; eliminating binding constraints leads to increased

---

<sup>50</sup> Nomothetic disciplines are systems of knowledge interested in arriving at general laws that govern human behavior based on systematically produced quantitative evidence through strictly defined and reproducible methods. They also aim to be space and context independent (Wallerstein, 1996).

private investment; and randomization produces internally valid impact estimates—hold true in all places at all times.

Like the way Shapin and Shaffer (2011) approached Hobbes' rationale rooted in natural philosophy, which Boyle and the experimentalists blithely dismissed, I employed a "charitable interpretation" of the arguments and claims of state experts that Treasury and MCC economists rejected. Such an interpretation is made possible by applying the principle of symmetry that neither subscribes to the rationality-politics divide so common in the planning and policy literature nor accepts the views of dominant experts regarding what is "rational" (or representative of collective social reason), and what is "irrational" (or representative of idiosyncratic interests). As a result, I argue that the characterizations that state economists made of other experts and their existing methods were inaccurate. Upon close empirical observation, the practices economists characterized as biased and subjective, either out of deceit or unawareness, appeared to be rooted in alternative rationalities.

In the case of governance indicators, foreign policy decisions were based largely on geopolitical calculation. Foreign policy community members wanted to direct funds to important US geopolitical allies such as Pakistan and Egypt. Most country government investment proposals either had strong place-based rationales or reflected the civil society community's concerns for marginalized populations in the periphery. While some countries presented laundry list proposals, the majority of countries based their proposals on targeted regional or demographic considerations even if the proposals spanned several sectors. Implementers selected farmers for success based on observable and unobservable criteria combined with theories of how agricultural innovations diffused. Implementers and field-level technicians were motivated by project success, which meant registering farmers that they

intuitively believed would excel. Meanwhile, the approaches of orthodox economists were not immune from bias; indeed, their methods favored certain interests and groups over others (e.g., relatively wealthier countries, urban development, and the average farmer).

The alternative rationalities in the cases were nevertheless vulnerable to economists' framing and standardized methods because they were more idiographic than nomothetic. Idiographic approaches to knowledge focus on the unique qualities of a particular unit (the prefix "idio" in ancient Greek refers to the distinct and personal) (Wallerstein, 1996).<sup>51</sup> This left those employing idiographic rationales vulnerable to criticisms that they were basing decisions on idiosyncratic (i.e., subjective) interests. Regarding governance indicators, while geopolitical realism is a social science sub-discipline, geopolitically oriented international realists believed that a country's unique bilateral relationship with the United States and particular geopolitical position in the world should inform funding decisions. With respect to growth diagnostics, development strategists holding the civil society tendency believed that a country region's or population's unique needs or assets should determine country investment strategies. As regards RCTs, field level technicians were free to deviate from observable criteria, such as years of education and wealth, that correlate strongly with "early adoption" in diffusion of innovation theory and select participants based on unique unobservable characteristics such as individuals' motivation and learning capacity. As a result, those who held idiographic rationales had a harder time arguing that they possessed a perspectival objectivity.

---

<sup>51</sup> Idiographic rationales are more deeply invested in case-study approaches that aim to understand an individual unit in its entirety than they are in establishing general laws about the behavior of a set of similar units. Context matters, and space and time are among some of its important elements. Evidence can be qualitative or quantitative but rarely precisely reproducible. History is the paradigmatic idiographic discipline (Wallerstein, 1996).

## **Overcoming Conflicting Rationalities**

After state economists arranged solutions by adapting CGD approved methods to solve organizational problems, fended off peers' alternative proposals, and successfully established themselves as spokespersons for objectivity, senior officials and economists became partners in method. Each actor, moreover, achieved some degree of autonomy within this partnership. Senior officials could use the methods to boost their department's legitimacy and by consequence bureaucratic autonomy. Economists could use their autonomy to stabilize their methods as governing rationales.

Scholars studying expert autonomy might be satisfied to stop here. After all, economists' power could now rest on both their specialized economic knowledge—drawn from their education and participation in epistemic communities—and autonomy from hierarchical superiors. Science studies scholars would point out, however, that state economists might still face trials of strength because other actors, especially other experts, opposed economists' methodological projects (Callon, 1984, 1986). In fact, state economists still had to enroll additional allies to overcome opposition from other expert such as the foreign policy community and project implementers that held alternative rationalities. Growth diagnostics faced the least resistance, while RCTs faced the most. The level of resistance to governance indicators fell somewhere in between. Consequently, each group of state economists—evaluation economists, Treasury economists, and MCC economists—had to exert a commensurate level of power.

**Trials of Strength in Global Development Finance.** Since expert claims routinely “affect, combat, refute, and negate” someone (Turner, 2007), it is not surprising that the claims of one expert community would conflict with those of another. Overcoming these conflicting rationalities was essential to the transformation of economic methods into governing rationales.

The cases here show that when orthodox economists at Treasury and the MCC sought to introduce and stabilize their methods, it led to a situation of conflicting rationalities—between their rationales and the rationales they sought to displace. When the now famous chemist Robert Boyle advanced his program of experimental science using vacuums, he had to contend with natural philosophers and their plenist ontologies (Shapin & Schaffer, 2011). Likewise, in the cases in this study, other experts and their associated rationalities had established operating norms that were present when economists promoted their rationales as governing principles. However, when conflicting rationalities are discussed in the planning and policy literature, it usually happens between state and non-state actors or between state formal rationality and non-state practical rationality (Goulet, 1986; Watson, 2003). Conflicting rationalities *within* the state among groups of bureaucrats or experts are rarely discussed.

In the case of growth diagnostics, it was sufficient for MCC economists to translate their superiors' interests and gain autonomy to stabilize growth diagnostics as a methodological approach to developing and designing compacts (see Table 7.2). Senior officials were looking for a way to streamline and discipline the compact development process and eliminate project “dogs” from consideration. Growth diagnostics were able to do this through a standardized, evidence-based process of elimination. There was no major competing rationality or expert network at the time of growth diagnostics' stabilization. Enrollment of senior officials was all that was necessary for growth diagnostics to become “obligatory passage points” for the resolution of development finance decisions (Callon, 1984).



Table 7.2

*Growth Diagnostics as Obligatory Passage Points*

Entity	Obstacle-problem	Entities' Goals	Why Translation Successful
MCC Compact Development Economists	Compact development and design using growth diagnostics	Conduct a growth diagnostic once a country is eligible for MCA funding to guide compact design	Not applicable
MCC Compact Development Senior Officials	"Laundry list" proposals; Laborious compact development negotiations	A "disciplined" and legitimate way to reject projects that were unworthy of funding  Avoid "political battles" with countries while receiving "technical cover" for decisions	Growth diagnostics would analyze a limited set of 10 constraints and used a standardized process to eliminate certain constraints from consideration

*Notes:* Table's framework adapted from Callon (1984).

However, in the other two cases in which state economists had to withstand trials of strength, they had to deploy various strategies of "power to" or power as persuasion rather than "power over" or power as coercion to enroll additional actors. State economists sought to transform what Callon (1984) calls bystanders and opponents into allies and make their methods obligatory passage points.

Treasury economists enrolled bystanders who shared economists' objectives such as congressional members and officials of the White House Office of Management and Budget (OMB) with the moral power of objectivity. While objectivity was important to economists in convincing their hierarchical superiors of the merits of their methods in all cases, it was particularly salient in the case of governance indicators. State economists' use of rules-based,

third party, quantitative methods and commitment to transparency made them appear impartial, selfless, and disciplined. These are all virtues of objectivity and contributed to a sense of economists' trustworthiness (Daston, 1992). Convincing others that one is objective can enhance a person's moral standing, which makes it a strategic form of persuasive power (Mehta & Winship, 2010). Once the moral standing that comes from objectivity in life and practice has been achieved, one can use the prestige earned as a resource to enroll others (Fuchs, 1997).

Having powerful congressional members and intellectual actors on their side was central to the success of Treasury economists in overcoming the resistance and alternative rationalities of the well-established foreign policy community. The interests of conservative congressional members and OMB officials were relatively easily translated, since they had their own reasons for wanting quantitative metrics established such as curbing development bureaucrats' discretion by increasing accountability mechanisms for foreign aid spending (see Table 7.3). Indicators enabled this because they provided metrics by which to target countries and monitor their progress. Treasury economists deployed a strategy of "trust in numbers" to enroll congressional representatives and their staff members and OMB officials. Porter (1996) described how bureaucratic agencies and their experts defensively adopted quantitative methods to fend off unwanted scrutiny and gain bureaucratic autonomy as a form of defense against unwanted intrusion. At certain moments in certain fields the public and its representatives trust agencies' quantitative methods more than officials' discretionary decision-making (Porter, 1996). While Porter's argument is very valuable, it nonetheless does not address the use of objectivity by experts as a form of *offense*, which as the case of governance indicators demonstrates, economists used against other expert groups such as the foreign policy community.

Table 7.3

*Governance Indicators as Obligatory Passage Points*

Entity	Obstacle-problem	Entities' Goals	Why Translation Successful
Treasury Economists	Selectivity by (governance) indicators	Build a transparent selection system that picks countries with the "best" policies and institutions conducive to generating growth	Not applicable
NSC International Economics Officials	Increased demands for global development finance; raising aid levels in a conservative administration and Republican-led congress	Raise aid levels to contribute to broader national security and growth agendas and meet the Monterrey Consensus' goal for greater development financing	Treasury economists' ran regression models that identified selection indicators that were positively correlated with growth
Republican-led Congress and OMB	Limited accountability for foreign aid which many believed was "failing"	Increase accountability mechanisms for foreign aid spending	Selection indicators provided a metric by which to monitor countries' progress. Increases in indicator scores, GDP, or preferably both would provide evidence of progress

Notes: Table's framework adapted from Callon (1984).

Evaluation economists enrolled opponents such as sector specialists, who resisted economists' objectives and whose interests were more challenging to translate with vehicular ideas and identity formation (see Table 7.4). In the case of RCTs, "impact" was an important vehicular idea. Vehicular ideas, similar to "boundary objects" (Star & Griesemer, 1989), emerge as ways of problem-solving and their significance can change with context. They are less ideas committed to some grand ideology that compels allegiance or rejection than inclusive, polysemic umbrellas under which a range of advocates can shelter, trade and shift their allegiances (Béland & Cox, 2016; McLennan, 2004). Evaluation economists, for example, could cast RCTs as a tool

for either accountability or learning. This was important to enrolling sector specialists in the enterprise of RCTs. MCC senior officials came to understand the impact of an MCC project in cost-benefit terms, or getting the “biggest bang for the MCC buck,” in the case of a particular intervention in another country. This same motivation, and thus conception of impact, was unsatisfactory to MCC sector specialists. Evaluation economists had to enroll sector specialists by casting impact as learning. Sector specialists were committed to and engaged in “best practice”. It became much more attractive for sector specialists to enroll in the impact evaluation efforts of evaluation economists if the impact represented opportunities to learn and improve project experiences and outcomes for participants (see Table 7.4).

Evaluation economists translated the interests of MCC sector specialists through their participation in evaluation workshops. The workshop served a similar purpose to that of the scientist’s laboratory (Latour, 1993; Law, 2011). The tactic of power used in evaluation economists’ workshops took the form of generosity rather than a monopoly of knowledge.<sup>52</sup> It was more important to evaluation economists that as many actors involved in project execution as possible understood—and more importantly, believed—in the merits of an RCT. After all, experts become more powerful and influential when they can graft their methods onto what others are doing (Rose, 1992).

Evaluation economists organized and conducted these workshops and invited academic economists to present and teach RCT principles. They shared the merits of RCTs widely and enthusiastically with those whose support was needed to successfully implement them. They framed the results of an RCT (i.e., project impact estimates) as educational outputs that served as

---

<sup>52</sup> Among those studying professions, monopolization over the supply of an expert’s knowledge is a major source of power (Eyal & Pok, 2015).

global public goods for development policy and practice at a time when economic research was increasingly being endorsed as an input to development alongside capital resources (Stiglitz, 1999). When evaluation economists were successful, they were able to cultivate new identities—co-experimenters. As such, sector specialists would engage in the righteous activity of generating educational public goods for the benefit of project participants globally rather than simply the pedestrian activity of delivering their specific project’s outputs.

Table 7.4

*Randomized Controlled Trials as Obligatory Passage Points*

Entity	Obstacle-problem	Entities’ Goals	Why Translation Successful
MCC Evaluation Economists	Identifying causal impacts of projects via randomized trials	Discover “what works” and replicate those interventions  Provide "unassailable" evidence of project impact	Not applicable
MCC Department of Accountability Senior Officials	Defining and operationalizing a results agenda to which MCC had publicly committed	Fulfilling MCC’s mission to “focus on results”	RCTs attributed impact to MCC investments, which was highly desirable for the accountability agenda
Sector Specialists	Avoiding project mistakes and failures of the past  Promoting and applying best practices	Project completion, improvement, and learning	Casted “impact” evaluation as learning rather than accountability  Translated sector specialists’ interests in workshops directed by outside evaluation experts

Notes: Table’s framework adapted from Callon (1984).

**The Resolution of Conflict and Rational-Legal Authority.** Science studies treats scientific work as political work because it is met with resistance. And wherever there is resistance, there is power (Foucault, 1978). These cases demonstrate that without postmodern forms of persuasive and productive power the power of experts' intellectual rationality could not be realized; their methods would neither have become obligatory passage points nor governing rationales. Hirschman and Berman (2014) suggest that it is important to understand whether economists' methods achieve the status of obligatory passage points. In all three of these cases, they did. However, their power was taken a step further in the cases of governance indicators and RCTs because once the enrollment of senior officials and other allies was successful and the methods of economists became obligatory passage points, they were codified in legally binding documents. Methods assumed the force of state rational-legal authority (Weber, 2009). Legislation establishing the MCA and MCC mandated that the MCC Board determine a candidate country's eligibility with objective and quantifiable indicators of a country's demonstrated commitment to "just and democratic governance" to the "maximum extent possible" (Pub. L. 108–199, 2004). Compacts' Monitoring and Evaluation Plans and contracts with implementers prescribed impact evaluation strategies and procedures, including randomization, which implementers had to follow if they wanted to remain compliant with contractual agreements.

There were no forms of rational-legal authority in the case of growth diagnostics. The need for binding language was unnecessary because there was no alternative rationality and expert group to overcome. Meanwhile, in the case of RCTs, rational-legal authority at the state organizational level was necessary but insufficient. Evaluation economists faced not one but two major sites of conflict—the state organization and the local implementing environment.

Consequently, commensurate forms of persuasive, postmodern forms of power were necessary to overcome implementers' continued opposition to randomized evaluation designs, in addition to the contractual agreement for randomization evaluation economists had already successfully drafted with sector specialists. Given the near daily incentives and pressures implementers faced to serve members of the control group, evaluation economists had to iteratively negotiate and persuade implementers to delay providing services to maintain the RCT's control group over the life of the project. Because implementers were not able to attend workshops, their appreciation of RCTs' merits and cultivation of identities had to be built over the course of multiple interactions. Implementers had to be reminded of the learning that was taking place and how everyone involved was engaged in the creation of educational public goods.

### **The Power of Methods**

Among those looking at experts' influence, many mainstream scholars in planning, policy, and political science have dismissed the influence of experts because it was limited to the instruments of planning and policy. Science studies scholars have taken experts' methods more seriously. They and others inspired by the field have shown us, through sustained scholarship over the past several decades, that methods enact the world around us. These scholars argue that methods should be thought of as interventions (Eyal & Levy, 2013). As such, they have important political effects. Such effects to, for example, make some things visible while obscuring others are described as small-*p* political effects and are both thoroughly documented elsewhere (Hirschman & Berman, 2014) and present in the cases studied here. In asking what effects methods have once stabilized, this study also raises the issue of methods' big-*P* political effects such as those on organizational power relations and the distribution of resources.

**Bureaucratic Power Relations.** Each successful trial of strength altered the nature of bureaucratic power relations. Through political activity and strategies of power, state economists gained not only autonomy but also *authority*. Senior officials allowed state economists to establish their favored economic method to make decisions on officials' behalf. In these cases, power was collective rather than zero-sum. According to Weberian "distributive" theories of power, one actor's power comes at the cost of another actor's power (Heiskala, 2001). However, the social theorist Talcott Parsons conceived of power as a collective force. Two actors can cooperate and enhance their joint power over third parties and nature. Actors are willing to sacrifice some of their own power if the collective power over a third party is large enough (Heiskala, 2001; Parsons, 1963).

In each of the cases presented here, senior officials and economists were willing to sacrifice some of their own autonomy and discretion by shifting them to the formal rules and procedures of economic methods. These situations were not so much a power struggle between management and economists, like the kind we may see between a sovereign and her subjects or a patriarch and his children, as it was a partnership based on method. Nonetheless, economists had relatively more to gain from this partnership—autonomy *and* the ability to withstand trials of strength against other expert groups to get their methods stabilized. As a result, the increased bureaucratic authority of state economists put them in charge of important decision-making functions. In turn, those who were dominant in decision-making were often relegated to operationalizing economists' methods. USAID and State Department foreign policy officials found themselves helping Treasury economists find appropriate indicators for the country selection system that Treasury championed. Sector specialists and implementers ensured that



evaluators' RCT designs would remain intact throughout implementation. Senior officials, by contrast, already had authority over these other experts.

**Decision Making.** Each of the three methods was deployed for a specific purpose. Governance indicators were vital to operationalizing selectivity for US bilateral aid effectiveness. To play this role, governance indicators transformed a country's political and institutional environment into a single number so that it could be compared against others. Growth diagnostics were central to efforts that sought to discipline the compact development process and narrow down country investment proposals. To achieve this they provided a taxonomy and process to organize a "clinical picture" of what ails a country's wellbeing (defined in terms of GDP) and quantified and classified a limited set of possible "cures." RCTs were fundamental to MCC becoming a pioneer in "managing for results" in development finance. To play this role, RCTs sorted "similar" persons into two spatially and temporally distinct groups, a treatment group that participates in a development project and receives its output and a control group for whom participation and outputs are denied or delayed.

Each method also represented its own unique form of power. To understand these forms of power, we had to first "get inside" each method by opening its black box (Latour, 1988; MacKenzie, 2005). Opening the black box of orthodox economic methods in the state revealed the various people and things such as survey respondents and project participant sample frames that state economists assembled to operationalize their methods. Economists introduced these actors and tools into the decision-making process.

Without the methods of the economists, much of the decision making at MCC resembled an "interactionist" form of planning between groups of politicians and experts that held both nomothetic and idiographic rationales. This was planning based on negotiation and "horse

trading”. In its ideal form, such a process would approximate what Charles Lindblom called “partisan mutual adjustment” where each actor pursued his or her interests based on information about what other actors were doing (Friedmann, 1987). However, the processes at MCC were similar to mutual adjustment processes more generally—they were laced with power relations and personal bias even while being informed by alternative rationales (Friedmann, 1987). These power relations and personal biases are part of the reason why economists denounced them. Nevertheless, partly supplanting partisan mutual adjustment with economic methods did not eliminate bias; it simply changed *whose* bias impacted decision-making.

Regarding governance indicators, decision making was shifted away from Washington, DC’s foreign policy community and towards an assemblage of country risk analysts, survey instruments, World Bank economists, and statistical models. “Some guy sitting in a café,” as one NSC official put it, that *The Economist Intelligence Unit* or similar publication purposively sampled to complete a survey would have his (or her) responses aggregated with hundreds of others into a single WGI indicator using a complex statistical model. This would determine a country’s governance rank. In the growth diagnostics case, diagnostics shifted compact development away from in-country government officials and sector experts towards MCC and core-team economists (and their consultants), a series of diagnostic tests, and global survey firms. MCC economists chose which constraints to consider, which to test, and the data and procedures that would be used to test them. This shaped the sectors and types of projects that could be considered for compact funding. With respect to RCTs, randomization displaced decision-making by sector specialists and project implementers with decisions made by evaluation economists (and their consultants), sample frames, and public lotteries. Evaluation economists identified projects amenable to RCTs and their consultants developed sample frames

of possible project participants. Randomization procedures (e.g., spreadsheet programs, lotteries) then sorted those in the sample frame into treatment and control groups.

These cases demonstrate that neither power nor bias can be avoided. Rather, the nature of power and bias changes depending on how planning and governance are approached. Replacing “interested” actors such as foreign policy officials, ministers of finance, and implementers engaged in negotiations and debate with economic methods meant introducing different entities and their respective interests into the decision-making process (see Table 7.5). These interests are not always explicit nor known to the creators or proponents of economic methods (Sinclair, 2008). For example, in the case of RCTs, the commitment of J-PAL economists to experimentation using the RCT made it virtually impossible to conduct evaluations of large-scale public goods such as reticulated infrastructure systems because of the SUTVA (stable-unit treatment value assumption) principle. When MCC evaluation economists adopted the RCT as an evaluation tool, they implicitly accepted the RCT’s inherent bias against networked public goods as part of the decisions about which projects will have a control group.

Table 7.5

*Influential Decision-makers Before and After Stabilization of Methods*

Pre-Method	Post-Method
Governance Indicators	
Senior State Department officials  Senior USAID officials	Treasury economists  Country-risk (and other) experts populating survey instruments  Governance survey and indicator producers  WGI's creators
Growth Diagnostics	
“Strong” ministry officials  MCC senior officials  Sector specialists  Members of the Poverty Reduction Strategy Process	MCC and core-team economists & consultants  Growth diagnostic creators  Growth diagnostics consultants  Respondents to competitiveness, enterprise, labor, and social surveys  MCC senior officials
RCTs	
MCC sector specialists  Project implementers  Local community leaders and officials	Evaluation economists  Stable Unit Treatment Value Assumption  Sample frames  Lottery systems

**The Political Effects of Methods.** Opening the black box also showed us how epistemological choices—such as correlating growth with governance or relying disproportionately on formal sector data—masked the political interests and values, or biases, of economists. These choices influenced each method’s outputs, such as governance rankings and diagnoses of binding constraints to growth. Understanding methods’ biases is important not only because they shape how we see and understand the world but because methods have material effects such as influencing the distribution of development finance resources.

Scholars have focused on experts' discourse and its ability to shape identities (Kendall & Wickham, 1999). But economists' methods also have distinct small-*p* political and big-*P* material effects (these are not always mutually exclusive). With respect to small-*p* political effects, governance indicators collapsed the incredible diversity of political and institutional environments across dozens of countries into a common metric. This created a ranked hierarchy among countries and made comparison of their institutions and polities easy. This disadvantaged Central Asian countries. They shared a unique institutional and political legacy that was different from regions that scored both well and poorly on the WGI. Furthermore, this scalar ranking created a zero-sum relationship and an "arms race" between countries, as one country's improvement in scores became automatically another country's loss.

The WGI also mobilized a theory of development (i.e., its political values) through its disproportionate reliance on investment risk analysts and indicator producers situated in the Global North. It advanced a lodestar model of state institutions based on emulation of countries such as Denmark. For countries in the Global South this meant fashioning state institutions whose *raison d'être* is the protection and enforcement of foreign capital's property rights for "market-enhancing" growth. I call this development model the "night watchman state," and heterodox economists have thoroughly documented the bias of governance indicator projects such as the WGI (Khan, 2007).

Through growth diagnostics economists linked economic theory with a data ecosystem that bolstered neoclassical economics and its adherents in the development planning process. This created a "regime of perceptibility" (Murphy, 2006) that focused MCC officials' attention on the diagnosis, or binding constraints, rather than a host of other historical, political, geographical, or cultural factors that might be responsible for the pathology of low private

investment. It also elevated this specific pathology over all others. A diagnosis provides structure to a disorderly narrative of symptoms and sorts the valid from the invalid (Jutel, 2011). Thus, growth diagnostics valorized countries' levels of private investment, which represents its creators' political values, while disregarding other issues such as high levels of inequality. They determined a pathology's cause by identifying and naming a disease, such as low social returns, based on a set of ten preexisting causes and a systematic process of reading symptoms and ruling out those causes. Like medical diagnoses, growth diagnostics prescribed the possible treatments, or investments and reforms in certain sectors to address underlying causes (Jutel, 2011). These processes were all conducted before the consultative process, thereby circumscribing civil society participation based on the diagnostic's results.

Randomized controlled trials experimented on, and thus subjected to scrutiny, a subset of development interventions that involved individualized goods or services such as training programs rather than public goods based on reticulated networks. Evaluation economists revealed what forms of knowledge they valued when they sought to actively exclude field-level technicians' intuitive and local knowledge of potential program participants as participation criteria. They believed that knowledge of project impact did not come from local praxis but quantifiable impact estimates based on observable characteristics and generalizable experimental methods.

Evaluation economists and their consultants extended their difference-in-difference matrix into the implementation landscape that transformed how implementers designed and executed these training projects. This "extension of the RCT rails" changed the temporal and spatial order in which implementers would normally deliver goods or services to project participants. This effect stemmed from evaluators having the power to determine who received

technical assistance and training first (treatment group) and who had to wait (control group) using random number generators and lotteries. Intervening in the implementation process in this way led to inconclusive or dampened project impact estimates and contributed to the sentiment that farmer-training projects “don’t work.”

In addition to the small-*p* political effects described above, methods in these cases also had material effects because they determined the distribution of resources. This is an understudied topic among scholars examining experts’ policy devices and the effects of expertise more generally. These big-*P* political effects were a consequence of economists’ epistemological choices that included the assumptions and intermediate decisions of method creators and state economists and the data they used in their methods. Regarding intermediate decisions, while Treasury economists did not directly choose countries, they chose the indicators that eventually chose the countries. They were in turn relying on Kaufmann and Kraay’s decisions about which survey datasets to include in their indicators. MCC economists did not choose sectors for investment, but they chose which constraints to test, which eventually identified sectors for investment. They built on the decisions of Hausmann, Rodrik, and Velasco about which constraints to include in their growth model. Evaluation economists chose which projects’ participants would be randomly assigned to a control group that eventually delayed benefits to a subset of poor farmers in the Agriculture and Rural Development Sector. They were following Duflo and Banerjee and CGD economists’ standards of evaluation evidence.

The data used in a method is also an important source of bias. Governance indicators rely on data that has a strong positive association with a country’s per capita GDP. Growth diagnostics rely on data that captures the views of larger, formal sector firms located in a country’s main cities rather than smaller, informal firms in small towns and rural areas.

Economists' intermediate decisions around how to conduct the methods and which data to use introduced systematic and structural, rather than idiosyncratic, biases into "rational" decision-making processes.

Together with the official decision-making authority methods possessed once stabilized, economists' epistemological choices had effects beyond those usually studied by other scholars. Governance indicators, for example, skewed the distribution of funding to countries that scored relatively higher on the WGI and were relatively wealthier in per-capita income terms. Because both perceptions of and actual changes to institutions are slow to change over time, it also meant repeatedly directing funds to a narrow group of countries. Growth diagnostics shifted funding to projects implemented in cities, particularly a country's capital city, rather than rural areas and away from construction and service delivery to policy and institutional reform. RCTs directed evaluation resources towards training projects in the Agriculture and Rural Development sector. They shifted resources in the form of technical assistance and agricultural inputs away from "pioneer" or "model" farmers to the "average farmer". While in this case RCTs actually equalized the distribution of resources, they also have the potential to shift resources away from the most needy individuals when they intervene in projects that target the poorest.

## **Conclusion**

The findings from these cases lead me to conclude that if experts do not exercise the kinds of power often considered as the reserve of politicians and similar actors with vested interests, the power of intellectual rationality cannot be realized. Furthermore, to compete in the state, a group of experts' particular intellectual rationality must contend with other rationalities, and, if it is to survive and thrive, overcome them. This conquest and survival is not something foretold—it is hard work. While state experts are increasingly becoming politicians, they are still



largely at the mercy of their hierarchical superiors and face path-dependent organizational norms (Centeno & Silva, 2016; Coats, 2001). To overcome these modern forms of power, they engage in institutional entrepreneurship through postmodern forms of “power as strategy” to enroll allies. Enrolling allies allows state experts to gain autonomy, overcome alternative rationalities, and establish their methods. It is often assumed that intellectual rationality exists “out there” as a perennial, naturally occurring phenomena just waiting to be applied in virgin organizational environments. This is a myth. Intellectual rationality must be constructed and nurtured (Cabantous et al., 2010). As a result, expertise is the work of stabilizing methods that can overcome conflicting rationalities and knowledge claims.

The cases have also led me to believe that understanding the power of state expertise requires bringing modern and postmodern theories of power together. Many studies rely either on Weberian analyses of authority, hierarchy, and resources or Foucauldian analyses of governmentality (e.g., Miller & Rose, 2008; Mills, 2000). It would be more productive, as some other scholars have pointed out, if Weber and Foucault were employed as supplements rather than as substitutes (Clegg et al., 2006; Flyvbjerg, 2004; Fraser, 1981). Moreover, the cases here point to the need to combine modern concepts of power such as hierarchy, authority, and resources with not only Foucauldian concepts and approaches but also science studies concepts of power such as translation, enrollment, and trials of strength (Callon, 1984; Law, 2011). This is especially so when investigating state experts and their methods.

Perhaps unpredictably, the use of these postmodern forms of strategic power led to new forms of modern, authoritative power. Foucauldian and other postmodernists depict experts’ methods as powerful but non-coercive (Rose & Miller, 1992). But when both politicians and experts transferred their decision-making authority to methods, and those methods became

codified in legal agreements after experts successfully prevailed in trials of strength, methods adopted the properties of official authority.

Finally, without looking at expert methods, we cannot understand the full extent of the power of intellectual rationality. Experts are powerful not only because they generate and spread ideas and discourse but because they establish methods that do work on their behalf. After becoming stabilized, economists' methods possessed not only the small-*p* political power to define, quantify, standardize, surveil, and normalize but also more modern forms of big-*P* political power to transform organizational power relations and determine the distribution of material resources.

## Chapter 8: Economic Expertise, Development Governance, and Planning Practice

### **Introduction**

For planning research, this dissertation's findings suggest a need to reconceptualize three common assumptions in planning scholarship using three concepts in science studies that have proven valuable to understanding state experts and their work. For development policy, the findings suggest how some of the methods studied affect participatory development and have implications for inter- and intra-country inequality. For planning practice, the results suggest how the study's approach to state economists' methods for governance could inspire reflective practice by economists and the allies they have enrolled in their methods. The study's findings also point to a possible path out of the "postmodern abyss" for practicing planners. If other experts could learn from the experiences of the orthodox economists discussed in this study, they might be able to address postmodernist concerns such as a rejection of dualisms and an emphasis on plurality and diversity. The chapter closes with a discussion of the methodological limitations of this study, and suggestions for future research.

### **Recap of the Study's Findings**

This dissertation posed questions to address shortcomings in the theories and research around the power of experts. First, I asked, how do state organizations come to adopt certain economic methods, or policy devices, as governing rationales? The cases demonstrate that in order to realize their specific entrepreneurial projects, which related to the stabilization of economic methods as policy devices for development governance, Treasury and Millennium Challenge Corporation (MCC) economic experts had to engage in political work—thought to be the reserve of politicians—by using various strategies of power. State economists overcame opposition to their proposed methods by mediating between their epistemic communities and

state organizations to achieve autonomy, presenting themselves as spokespersons for objectivity, and enrolling allies. Once the enrollment of senior officials and other allies was complete, state economists' methods became obligatory passage points.

The second question driving this study was, once adopted, how did these methods affect power relations, decision-making, and the distribution of resources? State economists' stabilization of methods put them in charge of important decision-making functions previously reserved for actors they had displaced. Their methods changed the decision-makers involved in development governance and introduced the ideas and values of a constellation of different actors to decision-making through economists' intermediate decisions and data choices. These new actors, values, and decisions introduced new biases and rationales that resulted in political effects. Consequently, Central Asian and other countries' unique circumstances could not be considered for decision-making and the poorest countries routinely failed to qualify for MCA funding; the length, breadth, and depth of the participatory planning process suffered and MCC projects began to resemble sectoral structural adjustment programs of the past; and projects delivering individual private goods were disproportionately selected for evaluation, participants thought more likely to succeed in those projects had their project outputs delayed, and the spatial and temporal sequencing of project implementation was altered.

### **Implications for Planning Research**

This study's findings have multiple implications for planning research, development governance, and planning practice. Here, the terms *planning* and *planners* refer broadly to fields dealing with public problems such as urban planning, public administration, development management, and public policy. This study suggests reconceptualizing three assumptions in planning research with heuristics for the social sciences. Reconceptualization heuristics are a

way of taking a familiar phenomenon and treating it as if it were an example of something quite different (Abbott, 2004). This dissertation's use of science studies' concepts such as the principle of symmetry, trials of strength, and opening the black box shows the value of reconceptualizing how planning scholarship addresses expert rationality, state experts, and expert methods, respectively.

The first common assumption in planning research is that experts' success or failure rests on the strength or weakness of their intellectual rationality. Too often, planning scholarship has made conclusions about expert influence based on the properties of intellectual rationality alone. Social reformers in the progressive era made normative assumptions about the value of intellectual rationality and its ameliorative effects on representative democracy and patronage politics (Finegold, 1995). This later led to an entire pedagogical and research program on rational and comprehensive planning (Altshuler, 1965b). More recently, empirical investigations of planning and practicing planners have cast doubt on the power of intellectual rationality (Baum, 1980; Flyvbjerg, 1998). John Forester is one of the few planning scholars who has presented a counterargument against claims of intellectual rationality's impotence by focusing on planners' communicative and discursive strategies for framing and agenda-setting in the face of power (Forester, 1988). The implicit assumption behind an emphasis on expert discourse is that if experts derive power from *speaking* truth to power, then it is worthwhile to either problematize expert speech or reform it to achieve progressive ends. However, focusing on the communicative rationality of experts delimits their sources of power and continues to put the "real" power of decision- and policy-making in the hands of non-experts (e.g., Fischer & Forester, 1993).

The power of experts, particularly through their strategies to gain autonomy and stabilize their methods, may be more extensive than originally thought. When examining state expertise, planning scholars should consider applying the principle of symmetry so as to provincialize rationality. Doing so treats rationality as a local accomplishment in particular contexts rather than a property that people or organizations do or do not possess (Cabantous et al., 2010). As a local accomplishment, agents such as experts must construct and perform rationality within institutions and organizations. Thus, the principle of symmetry suggests that when experts fail, they do not necessarily do so because rationality is a weak form of power but instead because they are ineffective politicians. The organizational theorist Jeffrey Pfeffer (1992) aptly said: “Knowledge without power is of remarkably little use. And power without the skill to employ it effectively is likely to be wasted” (Pfeffer, 1992: 342).

Some planning scholars in the 1980s (Baum, 1983b; Benveniste, 1989) began to explore this topic, but in recent decades this emphasis has faded, and the “blank space,” or cognitive blindness, between planners’ techniques and goals has reemerged (Baum, 1980). Planning research would benefit from scholars returning to experts’ sources and strategies of power as an explanation for planning and policy successes and failures.

The second common assumption in planning research is that conflicting rationalities exist between the state and its experts’ intellectual rationality and non-state actors and their practical rationality (Corburn, 2007; Goulet, 1986; Scott, 1998; Watson, 2003). This approach is based on an instrumentalized view of the state. Viewed instrumentally, the state has no internal conflicts or contradictions and lacks agents interested in or capable of conscious political action (Angotti & Marcuse, 2011). This conception of the state developed, for example, when a group of “neostatists” sought to dispel the structural theorists’ view of the state as “society-centric” (i.e., the

state as one among many outcomes of structural social forces) with the notion of state autonomy (Skocpol, 1985). In doing so, neostatists also inadvertently ended up casting the state as a monolithic entity with a unitary logic (Jessop, 2001). Weberian theories of bureaucracy that separate state and bureaucratic rationality from other forms of rationality, such as the practical rationality of politics or local communities, buttressed neo-statist claims.

In the planning and policy literature, conflicting rationalities *within* the state among groups of bureaucrats or experts are rarely discussed. The science studies concept of trials of strength reconceptualizes planning scholarship's view of conflicting rationalities. This study's findings suggest that the state is not monolithic with a unitary logic; it is an arena of conflicting rationalities. Economic experts faced opposition and had to withstand trials of strength from expert groups with alternative rationalities. If planning scholars applied the concept of trials of strength, they would look for instances where experts were opposed to each other and had to defend their work. They would find state environments rife with conflict and competing rationalities and would ask why certain groups opposed experts' proposals. Treating the state as an arena would make characterizations of the neoliberal state as "post-political" appear exaggerated. A particular version of the rational dream, such as neoliberalism, would be viewed as a political accomplishment. The dominance of economic rationality would be seen as orthodox economists withstanding trials of strength against other experts and rationalities. As the product of political work, these rationalities would be contingent, contestable, and in need of maintenance. There may not always be a trial of strength present, as was the case with growth diagnostics, but there will likely always be an alternative rationality. The issue is how dominant experts withstood trials of strength (if any occurred) or failed experts did not.

The third common assumption in planning research is that planning methods are ostensibly instruments of rationality. Planning scholarship presents planning methods as mostly technical instruments of rationality rather than political interventions. Opening the black box reveals their political and moral values. When planning scholars “get inside” methods, they help others to understand their political content (Alonso & Starr, 1989; Dorling & Simpson, 1999; Sinclair, 2008). Opening the black box in this way foreshadows methods’ possible political effects when planners use methods to bridge knowledge and action in planning environments. Each method is unique and therefore requires planning scholars to apply a tailored approach drawn from science studies, sociology, philosophy, history, and critical theory in order to uncover its political values and assumptions.

### **Implications for Development Governance**

Opening the black box of the three most popular economic methods in global development finance reveals various implications for development finance governance. Governance indicators, growth diagnostics, and RCTs have become staples of development governance at MCC and beyond. Multilateral development banks and various development finance agencies have adopted each of these methods with varying levels of enthusiasm. Growth diagnostics and RCTs have implications for participatory development, while governance indicators and growth diagnostics have implications for inequality.

Growth diagnostics and RCTs reduced the breadth and depth of participatory processes in development governance. Prior to growth diagnostics, compacts were based, to differing degrees, on countries’ poverty reduction strategy processes, or PRSPs. PRSPs emerged in the late 1990s in the wake of dissatisfaction with structural adjustment lending and policies. The consensus in the development community is that they are important yet imperfect vehicles for



participatory planning. However, one of the core missions of PRSPs was to maximize local civil-society participation in development planning. It included a participatory ideal for which to strive. In some cases, like PRSPs in Honduras, the extent of participation was genuine and significant (Dewachter & Molenaers, 2012). Planning through growth diagnostics moves away from this ideal and shifts some of the power over planning back to economic experts and development agencies' headquarters. This shift is a far cry from the days of structural adjustment planning when Washington-based economists' models doubled as universally valid prescriptions and absolute rules (Best, 2014), it nonetheless did forestall what could have been a deepening of the PRSP process, at least at MCC.

While PRSPs are meant to increase participation in the planning and design phase of development governance, participatory appraisal and similar constructivist and social-justice approaches to development learning emphasize participation in the project evaluation phase. In the 1980s, participation advocates won hard-fought battles to introduce these methods into existing policy analytic approaches. For example, Robert Chambers (Chambers, 1999) beseeched development practitioners to “put the last first” or listen to local peoples' analyses of their own problems and ideas for solutions. Participatory rural appraisals and action research linked advocacy planning and cooperative learning (Chambers, 1992). Evaluation was iterative, rather than a one-off exercise, and knowledge was co-produced through a variety of methods.

By contrast, in a randomized controlled trial knowledge is not co-produced with participants. Knowledge is collected through standardized survey instruments with predetermined categories and questions. The involvement of project participants in evaluation is limited to maintaining evaluators' research design, particularly the evaluation's control-group status. Participants must cooperate with the efforts of evaluators to minimize contamination and

spillover effects. Project participants were vessels into which interventions were poured and outcome data extracted rather than agents who interacted with projects in complex or contradictory ways. Their opinions about a project's services, inputs, or implementation quality were not solicited, because evaluators focused on attribution and accountability. Together, growth diagnostics and RCTs represent a reduction in the quantity and quality of participatory development planning and appraisal.

While growth diagnostics and RCTs affected participation, governance indicators and growth diagnostics had implications for development governance in another way: they both affected the distribution of development finance. Governance indicators had implications for inter-country inequality. Selecting countries based on governance indicators shifted development finance resources away from countries with the lowest per-capita gross national income. While some of these countries' incomes were low because of conflict and therefore they may not have made ideal candidates for large investment grants, some of them were considered simply too poorly governed to merit funding. Countries such as Comoros, Chad, Laos, Tajikistan, and Kyrgyzstan scored poorly on governance indicators and ranked among the poorest nations in the world. Consequently they did not qualify for MCA funding. Because institutions are path-dependent and resistant to change, countries with the lowest incomes failed to qualify year after year, while countries such as Cape Verde, Georgia, Ghana, and Morocco received funds repeatedly because they routinely passed the qualification criteria, which also included performance on human-development and economic-policy indicators. Over time, this has led to a curious conundrum in which MCC runs out of possible countries to fund while the focus on governance contributes to a "global underclass" of countries deemed needy but undeserving of development funding.

Growth diagnostics had implications for intra-country inequality. This diagnostic relies on neoclassical economic theory and formal sector data from enterprise and competitiveness surveys that together favored investment in metropolitan areas at the expense of rural areas. The framework of this diagnostic relies on a “root cause” approach toward constraints to private investment. This emphasized national-level policy and institutional reform over place-based “hard” investments in infrastructure and services and on making infrastructure services more affordable for existing users rather than on increasing access for new users. Prior to growth diagnostics, MCC investments were focused on integrated rural development programs, which had their own challenges and problems, and diverse transport investments that included sizeable programs for rural road construction. Compacts developed with diagnostics began to look like the World Bank’s sectoral structural-adjustment programs of the past and focused largely on institutional reform in country capitals.

In concert, governance indicators and growth diagnostics have channeled funding into national reform programs or urban investments in an increasingly select number of countries. Orthodox economists would likely find no fault with such a pattern. According to their theories of convergence, inter- and intra-country inequality will be reduced in the long term as prices approach equilibrium. Poorer countries and rural regions would eventually “catch up.” Reductions in regional inequality resemble an inverted U-curve, in which differences initially exacerbate but then decrease in later stages of development (Williamson, 1965). Economists, both orthodox and heterodox, and economic geographers have cast doubt on these theories and assumptions (Krugman, 1991; Myrdal, 1971; Pritchett, 1997). Without going into the detail of the alternative theories that challenge the neoclassical view, there is good reason to believe that cumulative causation—the idea that regional differences and urban and rural inequality are

exacerbated over time without state intervention to address inequality—is just as likely as convergence (Chakravorty, 2000).

Economic methods' effects on the distribution of development finance have consequences for economic governance and policy more broadly. In the wake of the 2008 global financial crisis, a debate has emerged over the role of orthodox economists in widening global inequality vis-à-vis other actors such as political elites, business leaders, and interests groups (Carrick-Hagenbarth & Epstein, 2012). While it was clear that inequality was on the rise, the role that economic expertise played in rising inequality appeared hazy (Galbraith, 2012). This study's conclusions suggest that orthodox economists play a role in exacerbating inequality. When orthodox economists gain autonomy and stabilize methods that relate strongly with aggregate measures of gross domestic product or rely upon neoclassical economic theory, their expertise aggravates inequality in the distribution of resources at least in the short to medium term.

### **Implications for Planning Practice**

This study has important implications for planning practice as well. Opening the black box of methods and understanding their political effects can prompt reflective practice. A practitioner engages in reflective practice when s/he critically assesses his or her own ways of thinking and doing things. In the case of global development finance, economists and their allies can ask themselves if the effects on participation and resource distribution are desirable for the targets of development finance. This would involve acknowledging the assumptions and limitations of their methods. At its best, reflective practice involves a sustained inquiry into the causes, meanings, and consequences of one's actions as part of a public debate (Fischler, 2012). Schon and his colleague Chris Argyris developed two models of learning—single loop and

double loop. Single-loop learning is the dominant mode of professional practice and is based on efficiency and the improvement of tasks, which are often technical (Argyris & Schon, 1992). The practitioner takes in new information and makes the appropriate adjustments to improve efficiency. Underlying norms, values, and frameworks remain unchanged. In contrast, double-loop learning redefines the ideas, perspectives, and norms that one brings to the performance of tasks (Argyris & Schon, 1992). Practitioners stuck in single-loop learning try to maximize control and minimize risk. Practitioners open to double-loop learning maximize the validity and transparency of information, including the values and objectives of actions (Fischler, 2012). They are willing to reveal political values and this sparks reflective practice.

Surprisingly, planners and other experts are not always aware of the political content and effects of their methods (Sinclair, 2008). State experts may possess enough skill to employ methods properly but rarely question the methods' underlying value-assumptions. They may miss the method's role in generating broader trends such as shifts in compact funding or biases in what programs to evaluate. Opening the black boxes of methods could reveal political values and effects to experts' allies. Usually, allies are removed even further from the inner workings of methods; they enrolled in the projects of experts because of their own interests or a partial and particular understanding of a method's purpose. Understanding the full scope of methods' effects might lead allies to reflect as well. Reflection on the part of both method proponents and enrolled allies could lead to reform, if so desired. This was the case at the Asian Development Bank. The Asian Development Bank's board of directors changed course when it learned that its country selection performance system, weighted heavily towards governance criteria, was favoring larger and wealthier countries at the expense of smaller and poorer ones. Or consider the mission statement of the United Kingdom-based *Radical Statistics Group*: "Radical Statistics

is a group of statisticians and others who share a common concern about the political assumptions implicit in the process of compiling and using statistics, and an awareness of the actual and potential misuse of statistics and its techniques” (Dorling & Simpson, 1999: xxiv).

Appetite for such reflection may not be as large as some might expect or like. Experts expend political capital to enroll allies and stabilize methods; openly questioning their methods’ assumptions and effects could appear too risky. Nonetheless, it still may be the case that recasting methods as political practice with political effects can highlight the individual agentic capacity of experts who may be what Ananya Roy calls “double agents” eager to reform their practices towards more progressive ends (Roy, 2012).

This study also offers one modest and admittedly partial route out of the postmodern abyss by characterizing the state as a plural arena of competing rationalities. The postmodern abyss represents a suspension between the decaying validity of the modernist project and convincing postmodernist critiques that are nevertheless discomfiting (Beauregard, 1991). Modernism emphasizes the centrality of reason and science to the human ability to act effectively via the state in a liberal democracy. Postmodernism questions the validity of the modern state and its experts’ claims to knowledge. Its intellectual foundations include a skepticism of conventional beliefs; rejection of universal bases for truth; discomfort with universal meanings attached to concepts; rejection of dualisms such as fact and value or rationality and irrationality; and belief in pluralism and a diversity of rationales that in the end may be irreconcilable (Harper & Stein, 1995). In response to the modernist project, postmodernists have enthusiastically embraced deconstructionism as a strategy of rigorous skepticism. However, the philosopher Hilary Putnam (1992) warned, “the philosophical irresponsibility of one decade can become the real-world political tragedy of a few decades later.

And deconstruction without reconstruction is irresponsibility” (Putnam, 1992: 132). Putnam as well as Latour (2004) have concluded that critique has “run out of steam” because a rejection of reason has left planning impotent and silent in the face of objective conditions of inequality, oppression, ignorance, and greed (Beauregard, 1991).

At the same time, postmodern accounts of knowledge, power, and expertise can become handmaidens to a program of “popular epistemology” that gives equal credit to all voices and forms of knowledge, including regressive and revanchist approaches that may either be indifferent or actively opposed to equality and ecosystem survival (DuPuis & Gareau, 2008; Latour, 2004). In the state, power enters not from the top down, but through the back door by means of particular well-funded interest groups.

Planning theorists have described three broad avenues that planners can follow to make a difference: work outside the state to affect planning and policy; withdraw from the state to create autonomous centers of governance and production; and operate within the state to make its characteristics more humane (Fainstein & Fainstein, 1979). The last option of working within the state is considered unfashionable and neglected in planning practice, often because it is not radical enough (Kraushaar, 1988). However, stabilizing alternative rationalities and methods within the state may offer progressive planners one of their best options for change. The examination of how certain experts gained power and how dominant methods came to be stabilized and have effects—as in the case of orthodox economists and orthodox economic rationality—raises questions about whether different (i.e., progressive) methods could produce different consequences.

Treating the state as an arena of competing rationalities between, for example, orthodox economists and other expert groups represents a “middle ground” that bridges the concerns of

postmodernists and engages in the kind of responsible reconstruction Putnam recommended. If the state is an arena of conflicting rationalities in which groups of experts deploy strategies to enroll others and fend off opposition to become obligatory passage points, then state organizations are institutions where plurality and diversity thrive; dualities between rationality and irrationality are rejected because power drives rationality; and rationalities are often irreconcilable because one group attempts to impose its rationality on organizational processes at the expense of others. This would only partially satisfy postmodernist concerns. When experts withstand trials of strength against other groups embracing other rationalities, they are not being antifoundationalist; instead, they are seeking to supplant one foundation of knowledge with another. Nor would these groups reject positivism or empiricism as the basis of knowledge and action. Some of the liberal tenets of modernism, such as the belief that (social) science can serve human needs, are maintained (Harper & Stein, 1995).

Paul Davidoff's plea for advocacy and pluralism in planning includes elements of this responsible reconstruction (Davidoff, 1965). Advocacy planning rejects master narratives and believes in plurality and diversity. It believes consensus is hard to reach and that the plans of different groups may be irreconcilable. What is needed instead is an honest broker to adjudicate between different plans. Peter Marris (1994) and Thomas Angotti (2011) constructively critique Davidoff's strategy by arguing that planners' approaches need to be more explicitly political and entrepreneurial. Leaving alternate plans to essentially speak for themselves before a jury does not suffice. Advocacy and equity planners and experts need to become "inside agitators" by entering the political arena and the state to make a case for their alternative rationalities (Angotti & Marcuse, 2011; Eisenstein, 1996; Marris, 1994). Consider, for example, what happened when feminists entered the Australian state in the late 1970s. These feminist bureaucrats, or



“femocrats,” significantly impacted the character of the Australian state, its budgets, and its public policies (Eisenstein, 1996). At the same time, their more radical counterparts in Britain, by contrast, rejected any association with the bureaucracy and likewise missed opportunities to bend the state towards feminist concerns (Watson, 1991). These approaches can extend to feminist (and other) methods as well (Bartlett, 1990; Nelson, 2001; Seitz & Adato, 2012; Roncolato, Reksten, & Grown, 2017).

If the state is an arena of conflicting rationalities, then feminists and other advocacy planners and experts can enter the state and stabilize progressive methods to make the state more humane (Sternberg, 1993). There are structural and institutional limits to the kind of expert methods that are possible in particular contexts and political moments (Garud et al., 2007; Innes, 1990). Nevertheless, within these limits there are important degrees of freedom and state experts can advance alternative rationalities through methods. Methods, after all, can promote equity and justice: “One could even argue that while economics is certainly consistent with legitimizing and masking many contemporary forms of domination ... this does not necessarily preclude the core [of applied economic practice] from playing less problematic roles ... a number of progressive economists have in fact tried to address major issues of gender domination, global inequality, and environmental degradation” (Reay, 2012: 78). While all forms of knowledge may be equally valid in a theoretical sense—and no group’s knowledge should be discounted as invalid—in a practical sense, decisions over budgets, regulations, policies, plans, and projects will inevitably rely on one set of experts and methods over another (Selinger & Crease, 2006; Turner, 2013). Progressive planners and experts could embrace the political work necessary to gain autonomy and withstand trials of strength to stabilize methods that, for example, address the commodification of land and labor to rehabilitate human habitation (Sternberg, 1993, 1996).

## **Study Limitations**

There is one main limitation to this study. The methods studied were observed in a relatively unique context. At the turn of the century, the MCA and MCC were established as a new US bilateral development finance program and agency, respectively, to reshape both US foreign assistance and global development finance. As a new agency, MCC faced fewer path dependencies and had more leeway to experiment with new approaches. It was also small. Its relatively compact size—no more than 300 employees globally between 2004 and 2015—made it a good candidate for a dissertation case study, but it also limits the representativeness of the study's findings. Other development finance contexts might produce different conclusions. For example, one interviewee remarked that MCC's use of growth diagnostics has diffused to the World Bank where the method faced initial resistance that was overcome. There was little to no resistance to growth diagnostics at MCC. Growth diagnostics have also spread to the US Agency for International Development (USAID), but the focus there has been on constraints to more inclusive forms of growth such as wages rather than private investment. At both the World Bank and USAID, growth diagnostics have been employed to inform country development strategies but, as of 2016, not to dictate which sectors should be targeted for investment. While the study's research design and approach controlled for this kind of variation with supplemental examples from multilateral development banks, additional cases would help corroborate whether MCC's experience with these methods is representative of the development finance field more broadly.

## **Suggestions for Future Research**

To address the limitations of this study, future research should focus on adding additional cases that vary according to the type of institutional context, method, and expert. These three

categories are not mutually exclusive and can be combined in different ways. First, scholars might investigate the nature of the economists' experiences with governance indicators, growth diagnostics, and RCTs in other organizations. For example, cases might explore these methods in multilateral development banks such as the World Bank or Inter-American Development Bank, non-profit organizations, and philanthropies that have been in existence for varying periods of time. Many of the dynamics described in this study have been anecdotally observed during the transition to structural adjustment in the World Bank during the 1980s (Pereira, 1995). That historical period could also be examined using this study's framework. If state economists are studied in additional contexts, we can begin to understand if the strategies orthodox economists used in this study are specific to the case of the MCA and MCC or more broadly applicable. The literature of science studies describes the processes of translation and enrollment broadly and focuses on scientists, engineers, and laboratories. This study described these processes for state economists and in considerable detail. The question that remains is whether these specific strategies have broader applicability in global development finance across organizational contexts.

It would also be helpful to understand whether these strategies hold among orthodox economists in other fields. Do economists working in public health or social work also become spokespersons of objectivity to gain power in their organizations? Do they draw on field-level standards from epistemic communities? Do they try to convince non-economists to join their cause by teaching them their craft? Or do orthodox economists in other fields use very different strategies to achieve autonomy and withstand trials of strength? It is equally important to understand the different methods of orthodox economists. Many economists draw on sub-disciplines in economics such as labor or environmental economics that use different methods.

A commonly used method not examined in this study is cost-benefit analysis. Economists also rely on models, which some scholars have argued are worth treating as methods when investigating their use as policy devices (e.g., Dellepiane-Avellaneda, 2015).

Finally, further research should focus on strategies by heterodox economists and other expert types. Do sociologists, environmental regulators, policy analysts, city planners and urban designers also achieve autonomy and withstand trials of strength? How do they do it? Do they use similar strategies to orthodox economists or are their strategies different? Are methods central to their political work within organizations? If stabilized, what effects do their methods have on their organizations and its outcomes? Research might focus on experts' attempts to stabilize "progressive" economic methods, like those that aim to highlight social and gender inequality, produce alternative measures of welfare, or account for ecosystem wealth. Studying various expert methods in diverse institutional contexts is important for a research program that emphasizes state experts' political work and their methods' political effects.

## Works Cited

### Scholarly and Secondary Sources

- 3ie. (2015). Impact Evaluation Repository. Retrieved April 23, 2018, from <http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/>
- Abbott, A. (1988). Transcending General Linear Reality. *Sociological Theory*, 6(2), 169–186. <https://doi.org/10.2307/202114>
- Abbott, A. (2004). *Methods of Discovery: Heuristics for the Social Sciences*. W.W. Norton & Company.
- Aguiar, L. L. M., & Schneider, C. J. (2016). *Researching Amongst Elites: Challenges and Opportunities in Studying Up*. Routledge.
- Akbulut, B., Adaman, F., & Madra, Y. M. (2015). The Decimation and Displacement of Development Economics. *Development and Change*, 46(4), 733–761. <https://doi.org/10.1111/dech.12181>
- Alonso, W., & Starr, P. (1989). *The Politics of Numbers*. Russell Sage Foundation.
- Altshuler, A. (1965a). Rationality and Influence in Public Service. *Public Administration Review*, 25(3), 226. <https://doi.org/10.2307/973747>
- Altshuler, A. (1965b). The Goals of Comprehensive Planning. *Journal of the American Institute of Planners*, 31(3), 186–195. <https://doi.org/10.1080/01944366508978165>
- Andersen, T. M., & Maibom, J. (2016). *The Big Trade-Off between Efficiency and Equity - Is it There?* (SSRN Scholarly Paper No. ID 2766484). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=2766484>
- Anderson, E. (2008). Experts, ideas, and policy change: the Russell Sage Foundation and small loan reform, 1909–1941. *Theory and Society*, 37(3), 271–310. <https://doi.org/10.1007/s11186-007-9050-0>
- Angotti, T., & Marcuse, P. (2011). *New York for Sale: Community Planning Confronts Global Real Estate*. MIT Press.
- Argyris, C., & Schon, D. A. (1992). *Theory in Practice: Increasing Professional Effectiveness*. Wiley.
- Arndt, C., & Oman, C. (2006). *Uses and abuses of governance indicators*. Paris: Development Centre of the Organisation for Economic Co-operation and Development.
- Asdal, K., Brenna, B., & Moser, I. (2007). *Technoscience: The Politics of Interventions*. Akademika Publishing.

- Babb, S. (2012). The Washington Consensus as transnational policy paradigm: Its origins, trajectory and likely successor. *Review of International Political Economy*, 20(2), 268–297. <https://doi.org/10.1080/09692290.2011.640435>
- Banerjee, A., Deaton, A., Lustig, N., Rogoff, K., & Hsu, E. (2006). *An Evaluation of World Bank Research, 1998 - 2005* (SSRN Scholarly Paper No. ID 2950327). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=2950327>
- Banerjee, A., & Duflo, E. (2008). *The Experimental Approach to Development Economics* (Working Paper No. 14467). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w14467>
- Banerjee, A., & Duflo, E. (2012). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. PublicAffairs.
- Banfield, E. C. (1961). *Political Influence*. Transaction Publishers.
- Barrett, C. B., & Carter, M. R. (2010). The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. *Applied Economic Perspectives and Policy*, 32(4), 515–548. <https://doi.org/10.1093/aep/ppq023>
- Bartlett, K. T. (1990). Feminist Legal Methods. *Harvard Law Review*, 103(4), 829–888. <https://doi.org/10.2307/1341478>
- Basu, K. (2005). New empirical development economics: remarks on its philosophical foundations. *Economic and Political Weekly*, 40(40), 4336–4339.
- Basu, K., & Stiglitz, J. E. (2016). *Inequality and Growth: Patterns and Policy: Volume I: Concepts and Analysis*. Springer.
- Battilana, J., Leca, B., & Boxenbaum, E. (2009). How actors change institutions : Towards a theory of institutional entrepreneurship. *The Academy of Management Annals*, Vol. 3, pp. 65–107. <https://doi.org/10.1080/19416520903053598>
- Baum, H. S. (1980). Analysts and Planners Must Think Organizationally. *Policy Analysis*, 6(4), 479–494.
- Baum, H. S. (1983a). Autonomy, Shame, and Doubt: Power in the Bureaucratic Lives of Planners. *Administration & Society*, 15(2), 147–184. <https://doi.org/10.1177/009539978301500201>
- Baum, H. S. (1983b). Politics and ambivalence in planners' practice. *Journal of Planning Education and Research*, 3(1), 13–22.
- Beamer, G. (2002). Elite Interviews and State Politics Research. *State Politics & Policy Quarterly*, 2(1), 86–96. <https://doi.org/10.1177/153244000200200106>

- Bearce, D. H., & Tirone, D. C. (2010). Foreign Aid Effectiveness and the Strategic Goals of Donor Governments. *The Journal of Politics*, 72(3), 837–851. <https://doi.org/10.1017/S0022381610000204>
- Beauregard, R. A. (1991). Without a Net: Modernist Planning and the Postmodern Abyss. *Journal of Planning Education and Research*, 10(3), 189–194. <https://doi.org/10.1177/0739456X9101000305>
- Béland, D. (2010). The Idea of Power and the Role of Ideas. *Political Studies Review*, 8(2), 145–154. <https://doi.org/10.1111/j.1478-9302.2009.00199.x>
- Béland, D., & Cox, R. H. (Eds.). (2010). *Ideas and Politics in Social Science Research*. Oxford University Press.
- Béland, D., & Cox, R. H. (2016). Ideas as coalition magnets: coalition building, policy entrepreneurs, and power relations. *Journal of European Public Policy*, 23(3), 428–445. <https://doi.org/10.1080/13501763.2015.1115533>
- Bell, D. (1965). *The End of Ideology* (Revised Edition edition). Free Press.
- Benveniste, G. (1972). *The Politics of Expertise*. Glendessary Press.
- Benveniste, G. (1989). *Mastering the politics of planning: crafting credible plans and policies that make a difference*. Jossey-Bass.
- Berman, E. P. (2011). *Creating the Market University: How Academic Science Became an Economic Engine*. Princeton University Press.
- Berndt, C. (2015). Behavioural economics, experimentalism and the marketization of development. *Economy and Society*, 44(4), 567–591. <https://doi.org/10.1080/03085147.2015.1043794>
- Bernstein, M. A. (2014). *A Perilous Progress: Economists and Public Purpose in Twentieth-Century America*. Princeton University Press.
- Berten, J., & Leisering, L. (2017). Social policy by numbers. How international organisations construct global policy proposals. *International Journal of Social Welfare*, 26(2).
- Besedovsky, N. (2018). Financialization as calculative practice: the rise of structured finance and the cultural and calculative transformation of credit rating agencies. *Socio-Economic Review*, 16(1), 61–84. <https://doi.org/10.1093/ser/mwx043>
- Best, J. (2014). *Governing Failure: Provisional Expertise and the Transformation of Global Development Finance*. Cambridge University Press.
- Beveridge, R. (2017). The (Ontological) Politics in Depoliticisation Debates: Three Lenses on the Decline of the Political. *Political Studies Review*, 15(4), 589–600. <https://doi.org/10.1177/1478929916664358>

- Beveridge, R., & Koch, P. (2017). The post-political trap? Reflections on politics, agency and the city. *Urban Studies*, 54(1), 31–43. <https://doi.org/10.1177/0042098016671477>
- Bezemer, D., & Headey, D. (2008). Agriculture, Development, and Urban Bias. *World Development*, 36(8), 1342–1364. <https://doi.org/10.1016/j.worlddev.2007.07.001>
- Bhatt, J. (2011). Causality and the Experimental Turn in Development Economics. *New School Economics Review*, 5(1), 50–64.
- Bimber, B. A. (1996). *The Politics of Expertise in Congress: The Rise and Fall of the Office of Technology Assessment*. SUNY Press.
- Birdsall, N., Levine, R., Lucas, S., & Shah, S. (2002). *On Eligibility Criteria for the Millennium Challenge Account*. Center For Global Development. Retrieved from <https://www.cgdev.org/publication/eligibility-criteria-millennium-challenge-account>
- Birdsall, N., Lucas, S., & Shah, S. (2002). *Guiding Principles for Design and Implementation of the MCA*. Center For Global Development. Retrieved from <https://www.cgdev.org/publication/guiding-principles-design-and-implementation-mca>
- Blackwill, A. R. D., & Harris, J. M. (2016). *War by Other Means: Geoeconomics and Statecraft*. Cambridge, Massachusetts: Belknap Press: An Imprint of Harvard University Press.
- Blackwill, R. D., & Harris, J. M. (2016). The Lost Art of Economic Statecraft: Restoring an American Tradition. *Foreign Aff.*, 95, 99.
- Blattman, C. (2008, March 1). Which is for you: MPA, MPA/ID, or PhD? Retrieved April 21, 2018, from <https://chrisblattman.com/2008/03/01/which-is-for-you-mpa-mpaid-or-phd/>
- Bloor, D. (1981). II.2 The Strengths of the Strong Programme. *Philosophy of the Social Sciences*, 11(2), 199–213.
- Bloor, D. (1991). *Knowledge and Social Imagery*. University of Chicago Press.
- Bøås, M., & McNeill, D. (2004). *Global Institutions and Development: Framing the World?* Psychology Press.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out classification and its consequences*. Cambridge, Mass.: MIT Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=13186>
- Boyer, M. C. (1986). *Dreaming the Rational City: The Myth of American City Planning*. MIT Press.
- Brainard, L. (2003). Compassionate conservatism confronts global poverty. *The Washington Quarterly*, 26(2), 147–169.



- Brainard, L., & Driscoll, A. (2003). *Making the Millennium Challenge Account Work for Africa* (Policy Brief No. 123). Brookings Institution. Retrieved from <https://www.brookings.edu/research/making-the-millennium-challenge-account-work-for-africa/>
- Breger, M. J., & Edles, G. J. (2000). Established by Practice: The Theory and Operation of Independent Federal Agencies. *Administrative Law Review*, 52, 1111.
- Breslau, D. (1997). The political power of research methods: Knowledge regimes in US labor-market policy. *Theory and Society*, 26(6), 869–902.
- Breslau, D. (1998). *In Search of the Unequivocal: The Political Economy of Measurement in U.S. Labor Market Policy*. Greenwood Publishing Group.
- Bretton Woods Project. (2003). *Poverty Reduction Strategy Papers (PRSPs): A Rough Guide*. Bretton Woods Project. Retrieved from <http://www.brettonwoodsproject.org/2003/04/art-16298/>
- Brint, S. (1990). Rethinking the policy influence of experts: From general characterizations to analysis of variation. In *Sociological Forum* (Vol. 5, pp. 361–385). Springer. Retrieved from <http://link.springer.com/article/10.1007/BF01115092>
- Broome, A., & Seabrooke, L. (2012). Seeing like an International Organisation. *New Political Economy*, 17(1), 1–16. <https://doi.org/10.1080/13563467.2011.569019>
- Brunsson, N., & Jacobsson, B. (2002). *A World of Standards* (Revised edition). Oxford: Oxford University Press.
- Burchell, G., Gordon, C., & Miller, P. (Eds.). (1991). *The Foucault Effect: Studies in Governmentality* (1 edition). Chicago: University Of Chicago Press.
- Burnside, A. C., & Dollar, D. (1997). Aid, policies, and growth. Retrieved from <https://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-1777>
- Burnside, C., Dollar, D., & others. (1998). *Aid, the incentive regime, and poverty reduction*. World Bank, Development Research Group, Macroeconomics and Growth. Retrieved from <https://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-1937>
- Bush, G. W. (2002). *The National Security Strategy of the United States of America*. Executive Office Of The President Washington DC. Retrieved from <http://www.dtic.mil/docs/citations/ADA406411>
- Bush, G. W., Hubbard, R. G., & Kroszner, R. S. (2003). *Economic Report of the President and the Annual Report of the Council of Economic Advisers, 2003*. Washington, DC: The Council of Economic Advisors.

- Cabantous, L., & Gond, J.-P. (2015). The Resistible Rise of Bayesian Thinking in Management: Historical Lessons From Decision Analysis. *Journal of Management*, 41(2), 441–470. <https://doi.org/10.1177/0149206314558092>
- Cabantous, L., Gond, J.-P., & Johnson-Cramer, M. (2010). Decision Theory as Practice: Crafting Rationality in Organizations. *Organization Studies*, 31(11), 1531–1566. <https://doi.org/10.1177/0170840610380804>
- Callon, M. (1984). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review*, 32, 196–233. <https://doi.org/10.1111/j.1467-954X.1984.tb00113.x>
- Callon, M. (1986). The Sociology of an Actor-Network: The Case of the Electric Vehicle. In *Mapping the Dynamics of Science and Technology* (pp. 19–34). Palgrave Macmillan, London. Retrieved from [https://link.springer.com/chapter/10.1007/978-1-349-07408-2\\_2](https://link.springer.com/chapter/10.1007/978-1-349-07408-2_2)
- Callon, M., & Law, J. (2005). On qualculation, agency, and otherness. *Environment and Planning D: Society and Space*, 23(5), 717–733. <https://doi.org/10.1068/d343t>
- Callon, M., & Rabeharisoa, V. (2003). Research “in the wild” and the shaping of new social identities. *Technology in Society*, 25(2), 193–204. [https://doi.org/10.1016/S0160-791X\(03\)00021-6](https://doi.org/10.1016/S0160-791X(03)00021-6)
- Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 8(1), 1–21. <https://doi.org/10.1080/19439342.2015.1034156>
- Carolini, G. Y. (2008). *Depreciating Development: How Accounting Under Fiscal Rigidity Imperils Slum Upgrading in São Paulo*. Columbia University.
- Carpenter, D. P. (2001). *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton University Press.
- Carrick-Hagenbarth, J., & Epstein, G. A. (2012). Dangerous interconnectedness: economists’ conflicts of interest, ideology and financial crisis. *Cambridge Journal of Economics*, 36(1), 43–63. <https://doi.org/10.1093/cje/ber036>
- Centeno, M. A. (1993). The new Leviathan: The dynamics and limits of technocracy. *Theory and Society*, 22(3), 307–335.
- Centeno, M. A. (2010). *Democracy Within Reason: Technocratic Revolution in Mexico*. Penn State Press.
- Centeno, M. A., & Silva, P. (2016). *The Politics of Expertise in Latin America*. Springer.
- Center for Global Development. (2014). *William Easterly vs. Owen Barder: The Tyranny of Experts*. Retrieved from <https://www.youtube.com/watch?v=p4QtWsNzO6c>

- Chakravorty, S. (2000). How does structural reform affect regional development? Resolving contradictory theory with evidence from India. *Economic Geography*, 76(4), 367–394.
- Chambers, R. (1992). *Rural Appraisal: Rapid, Relaxed and Participatory*. Institute of Development Studies.
- Chambers, R. (1999). *Whose Reality Counts?: Putting the First Last* (First Edition). Intermediate Technology Publications.
- Cheru, F. (2006). Building and Supporting PRSPs in Africa: What Has Worked Well so Far? What Needs Changing? *Third World Quarterly*, 27(2), 355–376.
- Chissano, J., Kufuor, J. A., & Mcpherson, P. (2002, July 5). The Right Way to Aid Africa. *Wall Street Journal*. Retrieved from <http://www.wsj.com/articles/SB1025827977315080080>
- Chwiero, J. M. (2007). Testing and measuring the role of ideas: The case of neoliberalism in the International Monetary Fund. *International Studies Quarterly*, 51(1), 5–30.
- Clegg, S. (1989). *Frameworks of Power*. SAGE Publications Ltd.
- Clegg, S. (1994). Weber and Foucault: Social Theory for the Study of Organizations. *Organization*, 1(1), 149–178. <https://doi.org/10.1177/135050849400100115>
- Clegg, S., Courpasson, D., & Phillips, N. (2006). *Power and Organizations*. SAGE.
- Coady, D., Grosh, M. E., & Hoddinott, J. (2004). *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. World Bank Publications.
- Coats, A. W. (2001). The Role of Economists in Government and International Agencies: A Fresh Look at the Field. *History of Economics Review*, 34(1), 19–32. <https://doi.org/10.1080/10370196.2001.11733356>
- Collingridge, D., & Reeve, C. (1986). *Science Speaks to Power: The Role of Experts in Policy Making*. St. Martin's Press.
- Collins, H., & Evans, R. (2009). *Rethinking Expertise*. University Of Chicago Press.
- Cook, N., Campbell, R. J., Brown, P., & Ratner, M. (2015, August 17). Powering Africa: Challenges of and U.S. Aid for Electrification in Africa [Report]. Retrieved April 21, 2018, from <https://digital.library.unt.edu/ark:/67531/metadc743408/>
- Cooper, F., & Packard, R. (Eds.). (1998). *International Development and the Social Sciences: Essays on the History and Politics of Knowledge* (1st ed.). University of California Press.
- Corburn, J. (2007). Community knowledge in environmental health science: co-producing policy expertise. *Environmental Science & Policy*, 10(2), 150–161. <https://doi.org/10.1016/j.envsci.2006.09.004>
- Crush, J. (2012). *Power of Development*. Routledge.

- Dargent, E. (2011). Agents or Actors? Assessing the Autonomy of Economic Technocrats in Colombia and Peru. *Comparative Politics*, 43(3), 313–332.
- Daston, L. (1992). Objectivity and the Escape from Perspective. *Social Studies of Science*, 22(4), 597–618.
- Daston, L. (1995). The moral economy of science. *Osiris*, 10, 2–24.
- Davidoff, P. (1965). Advocacy and Pluralism in Planning. *Journal of the American Institute of Planners*, 31(4), 331–338. <https://doi.org/10.1080/01944366508978187>
- Davig, T., & Leeper, E. M. (2007). Generalizing the Taylor Principle. *The American Economic Review*, 97(3), 607–635.
- Davis, K., Fisher, A., Kingsbury, B., & Merry, S. E. (2012). *Governance by Indicators: Global Power through Classification and Rankings*. Oxford University Press, USA.
- Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48(2), 424–455. <https://doi.org/10.1257/jel.48.2.424>
- Dellepiane-Avellaneda, S. (2015). The Political Power of Economic Ideas: The Case of “Expansionary Fiscal Contractions.” *The British Journal of Politics and International Relations*, 17(3), 391–418. <https://doi.org/10.1111/1467-856X.12038>
- Denzin, N. K. (1973). *The Research Act: A Theoretical Introduction to Sociological Methods*. Transaction Publishers.
- Desrosieres, A. (2001). How Real Are Statistics? Four Possible Attitudes. *Social Research*, 68(2), 339–355. <https://doi.org/10.2307/40971461>
- Desrosières, A. (2002). *The Politics of Large Numbers: A History of Statistical Reasoning*. Harvard University Press.
- Dewachter, S., & Molenaers, N. (2012). Who Takes a Seat at the Pro-Poor Table?: Civil Society Participation in the Honduran Poverty Reduction Strategy. *Latin American Research Review*, 46(3), 112–132. <https://doi.org/10.1353/lar.2011.0055>
- De Walque, D. de. (2013, June 18). Public randomization ceremonies [Text]. Retrieved April 21, 2018, from <http://blogs.worldbank.org/impactevaluations/public-randomization-ceremonies>
- DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147. <https://doi.org/10.2307/2095101>
- Dorling, D., & Simpson, S. (1999). *Statistics in Society: The Arithmetics of Politics*. Wiley.

- Dreher, A., Klasen, S., Vreeland, J. R., & Werker, E. (2013). The Costs of Favoritism: Is Politically Driven Aid Less Effective? *Economic Development and Cultural Change*, 62(1), 157–191. <https://doi.org/10.1086/671711>
- Drèze, J., & Stern, N. (1990). Policy reform, shadow prices, and market prices. *Journal of Public Economics*, 42(1), 1–45.
- Duflo, E. (1999). *Essays in empirical development economics* (Thesis). Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/9516>
- Duflo, E., Glennerster, R., & Kremer, M. (2007). *Using Randomization in Development Economics Research: A Toolkit* (SSRN Scholarly Paper No. ID 997109). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=997109>
- Dugger, C. W. (2007, December 7). U.S. Agency's Slow Pace Endangers Foreign Aid. *The New York Times*. Retrieved from <http://www.nytimes.com/2007/12/07/world/africa/07millennium.html>
- DuPuis, E. M., & Gareau, B. J. (2008). Neoliberal Knowledge: The Decline of Technocracy and the Weakening of the Montreal Protocol. *Social Science Quarterly*, 89(5), 1212–1229.
- Durlauf, S. N. (2009). The Rise and Fall of Cross-Country Growth Regressions. *History of Political Economy*, 41(Supplement 1), 315–333. <https://doi.org/10.1215/00182702-2009-030>
- Easterly, W. (2003). Can foreign aid buy growth? *The Journal of Economic Perspectives*, 17(3), 23–48.
- Easterly, W. (2007). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good* (Reprint edition). New York, NY: Penguin Books.
- Easterly, W. (2014). *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor*. New York: Basic Books.
- Eble, A., Boone, P., & Elbourne, D. (2017). On Minimizing the Risk of Bias in Randomized Controlled Trials in Economics. *The World Bank Economic Review*, 31(3), 687–707. <https://doi.org/10.1093/wber/lhw034>
- Ebrahim, A. S., & Rangan, V. K. (2009). *The Millennium Challenge Corporation and Ghana* (SSRN Scholarly Paper No. ID 1473215). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=1473215>
- Eisenstadt, S. N. (1958). Bureaucracy and Bureaucratization. *Current Sociology*, 7(2), 99–124. <https://doi.org/10.1177/001139215800700201>
- Eisenstein, H. (1996). *Inside Agitators: Australian Femocrats and the State*. Temple University Press.

- Ellwood, J. W., & Patashnik, E. M. (1993). In praise of pork. *Public Interest*, (110), 19.
- Engerman, D. C. (2013). Learning from the EastSoviet Experts and India in the Era of Competitive Coexistence. *Comparative Studies of South Asia, Africa and the Middle East*, 33(2), 227–238. <https://doi.org/10.1215/1089201X-2322507>
- Escobar, A. (2011). *Encountering Development: The Making and Unmaking of the Third World (New in Paper)*. Princeton University Press.
- Espeland, W. N. (1998). *The Struggle for Water: Politics, Rationality, and Identity in the American Southwest*. University of Chicago Press.
- Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113(1), 1–40. <https://doi.org/10.1086/517897>
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, 24, 313–343. <https://doi.org/10.2307/223484>
- Espeland, W. N., & Stevens, M. L. (2008). A Sociology of Quantification. *European Journal of Sociology / Archives Européennes de Sociologie*, 49(03), 401–436. <https://doi.org/10.1017/S0003975609000150>
- Essex, J. (2013). *Development, Security, and Aid: Geopolitics and Geoeconomics at the U.S. Agency for International Development* (1 edition). University of Georgia Press.
- Esteva, G., & Prakash, M. S. (1998). *Grassroots Post-Modernism: Remaking the Soil of Cultures*. Palgrave Macmillan.
- Eyal, G. (2013). For a Sociology of Expertise: The Social Origins of the Autism Epidemic. *American Journal of Sociology*, 118(4), 863–907. <https://doi.org/10.1086/668448>
- Eyal, G., & Buchholz, L. (2010). From the Sociology of Intellectuals to the Sociology of Interventions. *Annual Review of Sociology*, 36(1), 117–137. <https://doi.org/10.1146/annurev.soc.012809.102625>
- Eyal, G., & Levy, M. (2013). Economic Indicators as Public Interventions. *History of Political Economy*, 45(Supplement 1), 220–253. <https://doi.org/10.1215/00182702-2311007>
- Eyal, G., & Pok, G. (2015). What is security expertise? From the sociology of professions to the analysis of networks of expertise. In *Security Expertise: Practice, Power, Responsibility* (pp. 37–59). London: Routledge.
- Fainstein, N. I., & Fainstein, S. S. (1979). New debates in urban planning. *International Journal of Urban and Regional Research*, 3(1-3), 381–403. <https://doi.org/10.1111/j.1468-2427.1979.tb00796.x>

- Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing Practice and Practicing Theory. *Organization Science*, 22(5), 1240–1253. <https://doi.org/10.1287/orsc.1100.0612>
- Ferguson, J. (1994). *Anti-Politics Machine: Development, Depoliticization, and Bureaucratic Power in Lesotho*. University of Minnesota Press.
- Finegold, K. (1995). *Experts and Politicians: Reform Challenges to Machine Politics in New York, Cleveland, and Chicago*. Princeton University Press.
- Fischer, F., & Forester, J. (1993). *The Argumentative Turn in Policy Analysis and Planning*. Duke University Press.
- Fischler, R. (2012). Reflective practice. In B. Sanyal, L. J. Vale, & C. D. Rosan (Eds.), *Planning Ideas That Matter: Livability, Territoriality, Governance, and Reflective Practice* (pp. 313–332). Cambridge, Mass: The MIT Press.
- Flinders, M., & Wood, M. (2014). Depoliticisation, governance and the state. *Policy & Politics*, 42(2), 135–149. <https://doi.org/10.1332/030557312X655873>
- Flyvbjerg, B. (1996). *The Dark Side of Planning: Rationality and “Realrationalität”* (SSRN Scholarly Paper No. ID 2278431). Rochester, NY: Social Science Research Network. Retrieved from <http://papers.ssrn.com/abstract=2278431>
- Flyvbjerg, B. (1998). *Rationality and Power: Democracy in Practice*. (S. Sampson, Trans.). University Of Chicago Press.
- Flyvbjerg, B. (2001). *Making Social Science Matter: Why Social Inquiry Fails and How it Can Succeed Again*. (S. Sampson, Trans.) (1 edition). Oxford, UK; New York: Cambridge University Press.
- Flyvbjerg, B. (2004). Phronetic planning research: theoretical and methodological reflections. *Planning Theory & Practice*, 5(3), 283–306. <https://doi.org/10.1080/1464935042000250195>
- Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- Forester, J. (1988). *Planning in the Face of Power*. Berkeley: University of California Press.
- Foucault, M. (1978). *The History of Sexuality, Vol. 1: An Introduction* (Reissue edition). New York: Vintage Books.
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*. Pantheon Books.
- Foucault, M. (2001). *Power*. (J. D. Faubion, Ed., R. Hurley, Trans.) (1 edition). New York: The New Press.

- Foucault, M. (2012). *Discipline & Punish: The Birth of the Prison*. Random House Digital, Inc.
- Fox, J. A., & Brown, L. D. (Eds.). (1998). *The Struggle for Accountability: The World Bank, NGOs, and Grassroots Movements*. Cambridge, Mass: The MIT Press.
- Fraser, N. (1981). Foucault on modern power: Empirical insights and normative confusions. *Praxis International*, (3), 272–287.
- Freidson, E. (1988). *Professional Powers: A Study of the Institutionalization of Formal Knowledge* (1 edition). Chicago: University of Chicago Press.
- Friedmann, J. (1978). The epistemology of social practice. *Theory and Society*, 6(1), 75–92.
- Friedmann, J. (1987). *Planning in the Public Domain: From Knowledge to Action*. Princeton University Press.
- Fuchs, S. (1997). A Sociological Theory of Objectivity. *Science & Technology Studies*, 11(1), 4–26.
- Fukuda-Parr, S. (2011). Theory and Policy in International Development: Human Development and Capability Approach and the Millennium Development Goals. *International Studies Review*, 13(1), 122–132. <https://doi.org/10.1111/j.1468-2486.2010.01003.x>
- Fukuda-Parr, S., & Hulme, D. (2011). International Norm Dynamics and the “End of Poverty”: Understanding the Millennium Development Goals. *Global Governance: A Review of Multilateralism and International Organizations*, 17(1), 17–36. <https://doi.org/10.5555/ggov.2011.17.1.17>
- Galbraith, J. (1972). *The New Industrial State*. Oxford & IBH Publishing Company.
- Galbraith, J. (2012). *Inequality and Instability: A Study of the World Economy Just Before the Great Crisis*. Oxford University Press.
- Garud, R., Hardy, C., & Maguire, S. (2007). Institutional Entrepreneurship as Embedded Agency: An Introduction to the Special Issue. *Organization Studies*, 28(7), 957–969. <https://doi.org/10.1177/0170840607078958>
- Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company.
- Gieryn, T. F. (1983). Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review*, 48(6), 781–795. <https://doi.org/10.2307/2095325>
- Gilbert, C., & Vines, D. (2006). *The World Bank: Structure and Policies*. Cambridge University Press.



- Gilbert, F. (1984). *Machiavelli and Guicciardini: Politics and History in Sixteenth-century Florence*. Norton.
- Gilman, N. (2003). *Mandarins of the Future: Modernization Theory in Cold War America*. Johns Hopkins University Press.
- Glennerster, R., & Takavarasha, K. (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton University Press.
- Goldstein, M. (2011, November 29). Notes from the field: the usefulness of early workshops [Text]. Retrieved April 21, 2018, from <http://blogs.worldbank.org/impactevaluations/notes-from-the-field-the-usefulness-of-early-workshops>
- Gordon, D. M. (1963). Economic Advisory Group Receives Indonesian Grant. *The Harvard Crimson*. Retrieved from <http://www.thecrimson.com/article/1963/4/24/economic-advisory-group-receives-indonesian-grant/>
- Goulet, D. (1986). Three rationalities in development decision-making. *World Development*, 14(2), 301–317. [https://doi.org/10.1016/0305-750X\(86\)90061-6](https://doi.org/10.1016/0305-750X(86)90061-6)
- Government Accountability Office (1995). *Government Corporations: Profiles of Existing Government Corporations, December 1995*. GAO/GGD-96-14. GAO, Washington, DC.
- Grindle, M. S. (1977). Power, Expertise and the “Tecnico”: Suggestions from a Mexican Case Study. *The Journal of Politics*, 39(2), 399–426. <https://doi.org/10.2307/2130057>
- Gupta, A. (1998). *Postcolonial Developments: Agriculture in the Making of Modern India*. Duke University Press.
- Haas, P. M. (1992). Introduction: Epistemic Communities and International Policy Coordination. *International Organization*, 46(1), 1–35.
- Hall, P. (1989). *The Political Power of Economic Ideas: Keynesianism across Nations*. Princeton, N.J.: Princeton University Press.
- Hall, P. (2002). *Cities of Tomorrow: An Intellectual History of Urban Planning and Design in the Twentieth Century*. Wiley.
- Harper, T. L., & Stein, S. M. (1995). Out of the Postmodern Abyss: Preserving the Rationale for Liberal Planning. *Journal of Planning Education and Research*, 14(4), 233–244. <https://doi.org/10.1177/0739456X9501400401>
- Harsch, E. (2001). Africa preparing its own recovery plans. *Africa Recovery*, 15(1), 3–9.
- Hart, V. (1992). Feminism and bureaucracy: The minimum wage experiment in the District of Columbia. *Journal of American Studies*, 26(1), 1–22.

- Hattori, T. (2003). The moral politics of foreign aid. *Review of International Studies*, 29(2), 229–247.
- Hausmann, R., Klinger, B., & Wagner, R. (2008). *Doing Growth Diagnostics in Practice: A “Mindbook”* (CID Working Papers No. 177). Center for International Development at Harvard University. Retrieved from <https://ideas.repec.org/p/cid/wpfacu/177.html>
- Hausmann, R., Rodrik, D., & Velasco, A. (2004). Growth Diagnostics. Retrieved from <https://academiccommons.columbia.edu/catalog/ac:126975>
- Headey, D. (2009). Appraising a post-Washington paradigm: What Professor Rodrik means by policy reform. *Review of International Political Economy*, 16(4), 698–728. <https://doi.org/10.1080/09692290902740357>
- Heiskala, R. (2001). Theorizing power: Weber, Parsons, Foucault and neostructuralism. *Social Science Information*, 40(2), 241–264. <https://doi.org/10.1177/053901801040002003>
- Hellström, T. (2008). Transferability and Naturalistic Generalization: New Generalizability Concepts for Social Science or Old Wine in New Bottles? *Quality & Quantity*, 42(3), 321–337. <https://doi.org/10.1007/s11135-006-9048-0>
- Hirschman, A. O. (1981). *Essays in Trespassing: Economics to Politics and Beyond* (2 edition). Cambridge Eng. ; New York: Cambridge University Press.
- Hirschman, D. (2016). Stylized Facts in the Social Sciences. *Sociological Science*, 3, 604–626. <https://doi.org/10.15195/v3.a26>
- Hirschman, D., & Berman, E. P. (2014). Do economists make policies? On the political effects of economics. *Socio-Economic Review*. <https://doi.org/10.1093/ser/mwu017>
- Hood, K. (2017). The science of value: Economic expertise and the valuation of human life in US federal regulatory agencies. *Social Studies of Science*, 47(4), 441–465.
- Höpfl, H. M. (1999). Power, authority and legitimacy. *Human Resource Development International*, 2(3), 217–234. <https://doi.org/10.1080/136788699000000024>
- Hout, W. (2004). Political regimes and development assistance. *Critical Asian Studies*, 36(4), 591–613. <https://doi.org/10.1080/1467271042000273266>
- Hout, W. (2010). Governance and Development: changing EU policies. *Third World Quarterly*, 31(1), 1–12. <https://doi.org/10.1080/01436590903557298>
- Huber, J. D., & Shipan, C. R. (2002). *Deliberate Discretion?: The Institutional Foundations of Bureaucratic Autonomy*. Cambridge University Press.
- Ikenberry, G. J. (2009). Liberal Internationalism 3.0: America and the Dilemmas of Liberal World Order. *Perspectives on Politics*, 7(1), 71–87.

- Ikenberry, G. J. (2011). The future of the liberal world order: Internationalism after America. *Foreign Affairs*, 56–68.
- Iltan, S., & Phillips, L. (2010). Developmentalities and Calculative Practices: The Millennium Development Goals. *Antipode*, 42(4), 844–874. <https://doi.org/10.1111/j.1467-8330.2010.00778.x>
- Innes, J. E. (1990). *Knowledge and public policy: the search for meaningful indicators*. Transaction Publishers.
- International Finance Corporation. (2011). *Doing Business 2011 - Making a difference for entrepreneurs*. Retrieved from <http://www.doingbusiness.org/reports/global-reports/doing-business-2011>
- Jacobs, J. (1909). *Aesop. Fables. Vol. 17, Part 1. The Harvard Classics*. New York: P.F. Collier & Son.
- Jessop, B. (2001). Bringing the State Back In (Yet Again): Reviews, Revisions, Rejections, and Redirections. *International Review of Sociology*, 11(2), 149–173. <https://doi.org/10.1080/03906700020056029>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>
- Joyce, P. G. (2011). *The Congressional Budget Office: Honest Numbers, Power, and Policymaking* (Original edition). Washington D.C: Georgetown University Press.
- J-PAL. (2016). Evaluations | The Abdul Latif Jameel Poverty Action Lab. Retrieved April 23, 2018, from <https://www.povertyactionlab.org/evaluations>
- Justesen, L., & Mouritsen, J. (2011). Effects of actor - network theory in accounting research. *Accounting, Auditing & Accountability Journal*, 24(2), 161–193. <https://doi.org/10.1108/09513571111100672>
- Jutel, A. G. (2011). *Putting a Name to It: Diagnosis in Contemporary Society*. Johns Hopkins University Press.
- Kaja, A., & Werker, E. (2010). Corporate Governance at the World Bank and the Dilemma of Global Governance. *The World Bank Economic Review*, 24(2), 171–198. <https://doi.org/10.1093/wber/lhq006>
- Kamruzzaman, P. (2009). Poverty Reduction Strategy Papers and the rhetoric of participation. *Development in Practice*, 19(1), 61–71. <https://doi.org/10.1080/09614520802576385>
- Kanbur, R. (2001). Economic policy, distribution and poverty: the nature of disagreements. *World Development*, 29(6), 1083–1094.

- Kanbur, R. (2002). Economics, Social Science and Development. *World Development*, 30(3), 477–486. [https://doi.org/10.1016/S0305-750X\(01\)00117-6](https://doi.org/10.1016/S0305-750X(01)00117-6)
- Katz, M. B. (2013). *The Undeserving Poor: America's Enduring Confrontation with Poverty: Fully Updated and Revised*. OUP USA.
- Kaufmann, D., & Kraay, A. (2002). Governance indicators, aid allocation and Millennium Challenge Account. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=534063](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=534063)
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2), 220–246.
- Kaufmann, D., Kraay, A., & Zoido-Lobato, P. (2000). Governance matters: From measurement to action. *Finance and Development*, 37(2), 10.
- Kaufmann, D., Kraay, A., & Zoido, P. (1999a). *Aggregating Governance Indicators* (SSRN Scholarly Paper No. ID 188548). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=188548>
- Kaufmann, D., Kraay, A., & Zoido, P. (1999b). *Governance Matters* (SSRN Scholarly Paper No. ID 188568). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=188568>
- Kendall, G., & Wickham, G. M. (1999). *Using Foucault's Methods* (1st ed.). SAGE Publications Ltd.
- Khan, M. (2007). Governance, economic growth and development since the 1960s. Retrieved from [http://eprints.soas.ac.uk/9921/1/DESA\\_Governance\\_Economic\\_Growth\\_and\\_Development\\_since\\_1960s.pdf](http://eprints.soas.ac.uk/9921/1/DESA_Governance_Economic_Growth_and_Development_since_1960s.pdf)
- Khan, M. (2007). *Governance, Economic Growth and Development since the 1960s* (Working Paper). United Nations, Department of Economics and Social Affairs. Retrieved from <https://econpapers.repec.org/paper/unewpaper/54.htm>
- Kiely, R. (1999). The last refuge of the noble savage? A critical assessment of post - development theory. *The European Journal of Development Research*, 11(1), 30–55. <https://doi.org/10.1080/09578819908426726>
- Kilby, C. (2006). Donor influence in multilateral development banks: The case of the Asian Development Bank. *The Review of International Organizations*, 1(2), 173–195. <https://doi.org/10.1007/s11558-006-8343-9>
- Knack, S., & Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics*, 7(3), 207–227.

- Kosar, K. R. (2011). *Federal Government Corporations: An Overview*. Congressional Research Service.
- Kothari, U. (2005). Authority and expertise: The professionalisation of international development and the ordering of dissent. *Antipode*, 37(3), 425–446.
- Kraushaar, R. (1988). Outside the Whale: Progressive Planning and the Dilemmas of Radical Reform. *Journal of the American Planning Association*, 54(1), 91–100. <https://doi.org/10.1080/01944368808977159>
- Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3), 483–499. <https://doi.org/10.1086/261763>
- Lamont, M. (2012). Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology*, 38(1), 201–221. <https://doi.org/10.1146/annurev-soc-070308-120022>
- Larkin, B. (2013). The Politics and Poetics of Infrastructure. *Annual Review of Anthropology*, 42(1), 327–343. <https://doi.org/10.1146/annurev-anthro-092412-155522>
- Larson, M. S. (1979). *The Rise of Professionalism: A Sociological Analysis*. Berkeley u.a: Univ of California Pr.
- Latour, B. (1988). *Science in Action: How to Follow Scientists and Engineers Through Society* (Reprint edition). Cambridge, Mass.: Harvard University Press.
- Latour, B. (1993). *The Pasteurization of France*. Harvard University Press.
- Latour, B. (1998). *Prince for Machines as Well as for Machinations*. Edinburgh U. Press.
- Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, 30(2), 225–248.
- Latour, B. (2012). *We Have Never Been Modern*. Harvard University Press.
- Latour, B., & Woolgar, S. (2013). *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.
- Law, J. (2004). *After Method: Mess in Social Science Research* (1 edition). London ; New York: Routledge.
- Law, J. (2011). On power and its tactics: a view from the sociology of science. *The Sociological Review*, 34(1), 1–38.
- Leech, B. L. (2002). Asking Questions: Techniques for Semistructured Interviews. *PS: Political Science & Politics*, 35(4), 665–668. <https://doi.org/10.1017/S1049096502001129>
- Legovini, A. (2010). *Development impact evaluation initiative : a World Bank-wide strategic approach to enhance developmental effectiveness* (No. 63322) (pp. 1–72). The World

- Bank. Retrieved from <http://documents.worldbank.org/curated/en/689141468161083406/Development-impact-evaluation-initiative-a-World-Bank-wide-strategic-approach-to-enhance-developmental-effectiveness>
- Lilleker, D. G. (2003). Interviewing the political elite: Navigating a potential minefield. *Politics*, 23(3), 207–214.
- Lin, A. C. (1998). Bridging Positivist and Interpretivist Approaches to Qualitative Methods. *Policy Studies Journal*, 26(1), 162–180. <https://doi.org/10.1111/j.1541-0072.1998.tb01931.x>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE.
- Lindblom, C. E. (1979). *Usable Knowledge: Social Science and Social Problem Solving* (Later printing edition). New Haven: Yale University Press.
- Li, T. M. (2007). Governmentality. *Anthropologica*, 49(2), 275–281.
- Lucas, S. (2006a). Lessons from Ghana 3: Connecting Economic Growth and Poverty Reduction. Retrieved April 21, 2018, from <https://www.cgdev.org/page/lessons-ghana-3-connecting-economic-growth-and-poverty-reduction>
- Lucas, S. (2006b). What to Watch: Potential Challenges in Tanzania. Retrieved April 21, 2018, from <https://www.cgdev.org/page/what-watch-potential-challenges-tanzania>
- Lund, F. J., & Srinivas, S. (2000). *Learning from Experience: A Gendered Approach to Social Protection for Workers in the Informal Economy*. International Labour Organization.
- MacKenzie, D. (1981). *Statistics in Britain, 1865-1930: The Social Construction of Scientific Knowledge*. Edinburgh University Press.
- MacKenzie, D. (2005). Opening the black boxes of global finance. *Review of International Political Economy*, 12(4), 555–576. <https://doi.org/10.1080/09692290500240222>
- MacKenzie, D., Muniesa, F., & Siu, L. (2007). *Do Economists Make Markets?: On the Performativity of Economics*. Princeton University Press.
- Maiers, C. (2017). Analytics in action: users and predictive data in the neonatal intensive care unit. *Information, Communication & Society*, 20(6), 915–929. <https://doi.org/10.1080/1369118X.2017.1291701>
- Malik, A. (2002). *State of the Art in Governance Indicators* (Human Development Occasional Papers (1992-2007) No. HDOCPA-2002-07). Human Development Report Office (HDRO), United Nations Development Programme (UNDP). Retrieved from <https://ideas.repec.org/p/hdr/hdocpa/hdocpa-2002-07.html>

- Mamadouh, V. D. (1998). Geopolitics in the nineties: one flag, many meanings. *GeoJournal*, 46(4), 237–253.
- Mannheim, K. (1950). *Freedom, power, and democratic planning*. Oxford University Press.
- Mansnerus, E. (2013). Using model-based evidence in the governance of pandemics. *Sociology of Health & Illness*, 35(2), 280–291. <https://doi.org/10.1111/j.1467-9566.2012.01540.x>
- Marcella, G. (2008). *Affairs of State: The Interagency and National Security*. The Army War College Strategic Studies Institute. Retrieved from <http://www.dtic.mil/docs/citations/ADA491602>
- Markoff, J., & Montecinos, V. (1993). The ubiquitous rise of economists. *Journal of Public Policy*, 13, 37–37.
- Marris, P. (1994). Advocacy Planning As a Bridge Between the Professional and the Political. *Journal of the American Planning Association*, 60(2), 143–146. <https://doi.org/10.1080/01944369408975563>
- Mathison, S. (1988). Why Triangulate? *Educational Researcher*, 17(2), 13–17. <https://doi.org/10.3102/0013189X017002013>
- Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics*, 110(3), 681–712. <https://doi.org/10.2307/2946696>
- Maxwell, S., & Stone, D. D. L. (Eds.). (2004). *Global Knowledge Networks and International Development*. Routledge.
- McCann, P., & Rodríguez-Pose, A. (2011). Why and When Development Policy Should Be Place-Based. In *OECD Regional Outlook 2011* (pp. 203–213). Organisation for Economic Co-operation and Development. Retrieved from <http://www.oecd-ilibrary.org/content/chapter/9789264120983-16-en>
- McGrattan, E. R., & Schmitz, J. A. (1999). Chapter 10 Explaining cross-country income differences. *Handbook of Macroeconomics*, 1, 669–737. [https://doi.org/10.1016/S1574-0048\(99\)01013-7](https://doi.org/10.1016/S1574-0048(99)01013-7)
- McKenzie, D. (2015, March 30). Tips for Randomization in the Wild: Adding a Waitlist [Text]. Retrieved April 23, 2018, from <http://blogs.worldbank.org/impacetevaluations/tips-randomization-wild-adding-waitlist>
- McLennan, G. (2004). Travelling with vehicular ideas: The case of the third Way. *Economy and Society*, 33(4), 484–499. <https://doi.org/10.1080/0308514042000285251>
- Mehta, J. (2009, April). *The Chastened Dream : Knowledge, Action and Professional Education*. Power Point presented at the Civic and Moral Education Initiative Colloquium, Harvard University. Retrieved from <http://ssc.bibalex.org/viewer/detail.jsf?lid=9BD726324DC7124F00D16C40357CFECD>

- Mehta, J. (2013). How Paradigms Create Politics: The Transformation of American Educational Policy, 1980-2001. *American Educational Research Journal*, 50(2), 285–324. <https://doi.org/10.3102/0002831212471417>
- Mehta, J., & Winship, C. (2010). Moral power. In *Handbook of the Sociology of Morality* (pp. 425–438). Springer. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4419-6896-8\\_22](http://link.springer.com/chapter/10.1007/978-1-4419-6896-8_22)
- Milberg, W. (2009). The New Social Science Imperialism and the Problem of Knowledge in Contemporary Economics. *Economic Persuasions*, 3, 43.
- Miller, P., & Rose, N. (2008). *Governing the Present: Administering Economic, Social and Personal Life*. Polity.
- Mills, C. W. (2000). *The Power Elite*. Oxford University Press.
- Mills, S. (2012). *Michel Foucault*. Routledge.
- Mintrom, M. (1997). Policy Entrepreneurs and the Diffusion of Innovation. *American Journal of Political Science*, 41(3), 738–770. <https://doi.org/10.2307/2111674>
- Mitchell, T. (2002). *Rule of Experts: Egypt, Techno-Politics, Modernity* (1st ed.). University of California Press.
- Mitchell, T. (2005). The work of economics: how a discipline makes its world. *European Journal of Sociology / Archives Européennes de Sociologie*, 46(02), 297–320. <https://doi.org/10.1017/S000397560500010X>
- Mookherjee, D. (2005). Is There Too Little Theory in Development Economics Today? *Economic and Political Weekly*, 40(40), 4328–4333.
- Morgenthau, H. J. (1963). *Politics Among Nations: The Struggle for Power and Peace*. Knopf.
- Morris, S. (2017). Time for Some Credit Counseling for the World Bank’s Biggest Deadbeat Donor. Retrieved April 21, 2018, from <https://www.cgdev.org/blog/time-some-credit-counseling-world-banks-biggest-deadbeat-donor>
- Moynihan, D. P. (2009). Through A Glass, Darkly: Understanding the Effects of Performance Regimes. *Public Performance & Management Review*, 32(4), 592–603. <https://doi.org/10.2753/PMR1530-9576320409>
- Mügge, D. (2016). Studying macroeconomic indicators as powerful ideas. *Journal of European Public Policy*, 23(3), 410–427. <https://doi.org/10.1080/13501763.2015.1115537>
- Murphy, M. (2006). *Sick Building Syndrome and the Problem of Uncertainty: Environmental Politics, Technoscience, and Women Workers* (1 edition). Durham N.C.: Duke University Press.



- Myrdal, G. (1971). *Economic theory and underdeveloped regions* (First Thus edition). New York: Harper & Row.
- Nader, L. (1969). Up the Anthropologist: Perspectives Gained From Studying Up. In *Reinventing Anthropology* (pp. 284–311). New York: Random House.
- Nathan, R. P. (1987). The Politics of Printouts: the use of official numbers to allocate federal grants-in-aid. In *The Politics of Numbers* (pp. 331–342). New York: Russell Sage Foundation.
- Nelkin, D. (1975). The Political Impact of Technical Expertise. *Social Studies of Science*, 5(1), 35–54.
- Nelson, J. A. (2001). Economic methodology and feminist critiques. *Journal of Economic Methodology*, 8(1), 93–97. <https://doi.org/10.1080/13501780010023252>
- Nobel Media. (2014). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1993. Retrieved April 23, 2018, from [https://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/1993/](https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1993/)
- North, D. (1966). *The Economic Growth of the United States: 1790-1860* (First Edition). New York: W. W. Norton & Company.
- North, D. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.
- North, D. (1994). Economic Performance Through Time. *The American Economic Review*, 84(3), 359–368.
- North, D. (1995). The new institutional economics and third world development. In *The New Institutional Economics and Third World Development* (Vol. 21). Routledge, London and New York.
- North, D., & Thomas, R. P. (1976). *The Rise of the Western World: A New Economic History*. Cambridge University Press.
- Oakeshott, M. (1991). *Rationalism in Politics and Other Essays*. LibertyPress.
- Okun, A. M. (1975). *Equality and Efficiency: The Big Tradeoff*. Washington: Brookings Institution Press.
- Osborne, T. (2004). On mediators: Intellectuals and the ideas trade in the knowledge society. *Economy and Society*, 33(4), 430–447. <https://doi.org/10.1080/0308514042000285224>
- Pankaj, A. K. (2005). Revisiting Foreign Aid Theories. *International Studies*, 42(2), 103–121. <https://doi.org/10.1177/002088170404200201>
- Parsons, C. (2002). Showing Ideas as Causes: The Origins of the European Union. *International Organization*, 56(1), 47–84. <https://doi.org/10.1162/002081802753485133>

- Parsons, T. (1963). On the concept of political power. *Proceedings of the American Philosophical Society*, 107(3), 232–262.
- Peattie, L. (1983). Realistic planning and qualitative research. *Habitat International*, 7(5), 227–234. [https://doi.org/10.1016/0197-3975\(83\)90073-5](https://doi.org/10.1016/0197-3975(83)90073-5)
- Pereira, L. C. B. (1995). Development economics and the World Bank's identity crisis. *Review of International Political Economy*, 2(2), 211–247. <https://doi.org/10.1080/09692299508434318>
- Peters, B. G., & Pierre, J. (1998). Governance Without Government? Rethinking Public Administration. *Journal of Public Administration Research and Theory*, 8(2), 223–243. <https://doi.org/10.1093/oxfordjournals.jpart.a024379>
- Petty, J., & Heimer, C. A. (2011). Extending the rails: How research reshapes clinics. *Social Studies of Science*, 41(3), 337–360. <https://doi.org/10.1177/0306312710396402>
- Pfeffer, J. (1992). *Managing with Power: Politics and Influence in Organizations*. Harvard Business Press.
- Pfiffner, J. P. (2007). The first MBA president: George W. Bush as public administrator. *Public Administration Review*, 67(1), 6–20.
- Phillips, D., Waddington, H., & White, H. (2014). Better targeting of farmers as a channel for poverty reduction: a systematic review of Farmer Field Schools targeting. *Development Studies Research*, 1(1), 113–136. <https://doi.org/10.1080/21665095.2014.924841>
- Pickering, A. (1992). *Science as Practice and Culture*. University of Chicago Press.
- Pinch, T. J. (1992). Opening Black Boxes: Science, Technology and Society. *Social Studies of Science*, 22(3), 487–510. <https://doi.org/10.1177/0306312792022003003>
- Piron, L.-H., & Evans, A. (2004). *Politics and the PRSP approach: synthesis paper*. London: Overseas Development Institute.
- Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. Clarendon Press.
- Porter, D., & Craig, D. (2004). The third way and the third world: poverty reduction and social inclusion in the rise of “inclusive” liberalism. *Review of International Political Economy*, 11(2), 387–423.
- Porter, T. M. (1996). *Trust in Numbers*. Princeton University Press.
- Power, M. (2004). Counting, Control and Calculation: Reflections on Measuring and Management. *Human Relations*, 57(6), 765–783. <https://doi.org/10.1177/0018726704044955>

- Pritchett, L. (1997). Divergence, Big Time. *Journal of Economic Perspectives*, 11(3), 3–17. <https://doi.org/10.1257/jep.11.3.3>
- Pritchett, L. (2002). It pays to be ignorant: A simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, 5(4), 251–269. <https://doi.org/10.1080/1384128032000096832>
- Putnam, H. (1992). *Renewing philosophy*. Harvard University Press.
- Radelet, S. (2003). *Challenging Foreign Aid: A Policymaker's Guide to the Millennium Challenge Account*. Washington, DC: Center for Global Development.
- Rao, H. (1998). Caveat Emptor: The Construction of Nonprofit Consumer Watchdog Organizations. *American Journal of Sociology*, 103(4), 912–961. <https://doi.org/10.1086/231293>
- Ravallion, M. (2012). Mashup Indices of Development. *The World Bank Research Observer*, 27(1), 1–32. <https://doi.org/10.1093/wbro/lkr009>
- Reay, M. (2007). Academic Knowledge and Expert Authority in American Economics. *Sociological Perspectives*, 50(1), 101–129. <https://doi.org/10.1525/sop.2007.50.1.101>
- Reay, M. J. (2012). The Flexible Unity of Economics. *American Journal of Sociology*, 118(1), 45–87. <https://doi.org/10.1086/666472>
- Reddy, S. (2012). Randomise This! On Poor Economics. *Review of Agrarian Studies*, 2(2), 60–73.
- Reed, M. I. (2012). Masters of the universe: Power and elites in organization studies. *Organization Studies*, 33(2), 203–221.
- Rodrik, D. (2008). *The New Development Economics: We Shall Experiment, but How Shall We Learn?* (SSRN Scholarly Paper No. ID 1296115). Rochester, NY: Social Science Research Network. Retrieved from <http://papers.ssrn.com/abstract=1296115>
- Rodrik, D. (2009). *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*. Princeton: Princeton University Press.
- Rogers, E. M. (1962). *Diffusion of innovations*. Free Press of Glencoe.
- Roncolato, L., Reksten, N., & Grown, C. (2017). Engendering Growth Diagnostics: Examining Constraints to Private Investment and Entrepreneurship. *Development Policy Review*, 35(2), 263–287. <https://doi.org/10.1111/dpr.12208>
- Rose, N. (1992). Engineering the Human Soul: Analyzing Psychological Expertise. *Science in Context*, 5(2), 351–369. <https://doi.org/10.1017/S0269889700001228>

- Rose, N., & Miller, P. (1992). Political Power beyond the State: Problematics of Government. *The British Journal of Sociology*, 43(2), 173–205. <https://doi.org/10.2307/591464>
- Rose, S., Kalow, J., Parks, B. C., & Masaki, T. (2016). *The Use and Utility of US Government Approaches to Country Ownership: New Insights from Partner Countries*. Washington, D.C.: Center For Global Development. Retrieved from <https://www.cgdev.org/publication/use-and-utility-us-government-approaches-country-ownership>
- Rose, S., & Wiebe, F. (2015a). *Focus on Policy Performance: MCC's Model in Practice*. Center For Global Development. Retrieved from <https://www.cgdev.org/publication/focus-policy-performance-mccs-model-practice>
- Rose, S., & Wiebe, F. (2015b). *Focus on Results: MCC's Model in Practice*. Center For Global Development. Retrieved from <https://www.cgdev.org/publication/focus-results-mccs-model-practice>
- Roy, A. (2012). Ethnographic Circulations: Space–Time Relations in the Worlds of Poverty Management. *Environment and Planning A*, 44(1), 31–41. <https://doi.org/10.1068/a44180>
- Ruckert, A. (2009). A Decade of Poverty Reduction Strategies in Latin America: Empowering or Disciplining the Poor? *Labour, Capital and Society / Travail, Capital et Société*, 42(1/2), 56–81.
- Sachs, J. (2006). *The End of Poverty: Economic Possibilities for Our Time*. Penguin.
- Sachs, W. (1992). *The Development Dictionary: A Guide to Knowledge as Power*. Zed Books.
- Sampson, R. J. (2010). Gold Standard Myths: Observations on the Experimental Turn in Quantitative Criminology. *Journal of Quantitative Criminology*, 26(4), 489–500. <https://doi.org/10.1007/s10940-010-9117-3>
- Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, 18(2), 179–183. <https://doi.org/10.1002/nur.4770180211>
- Sanger, D. E. (2001, January 16). Bush Plans to Stress Effects Of Economics on Security. *The New York Times*. Retrieved from <https://www.nytimes.com/2001/01/16/world/bush-plans-to-stress-effects-of-economics-on-security.html>
- Savedoff, W. D., & Levine, R. (2006). Learning from Development: the Case for an International Council to Catalyze Independent Impact Evaluations of Social Sector Interventions. Retrieved April 21, 2018, from <https://www.cgdev.org/publication/learning-development-case-international-council-catalyze-independent-impact-evaluations>
- Scholz, R. W., & Tietje, O. (2002). *Embedded Case Study Methods: Integrating Quantitative and Qualitative Knowledge*. SAGE.

- Schön, D. A., Drake, W. D., & Miller, R. I. (1984). Social Experimentation as Reflection-in-Action Community-Level Nutrition Intervention Revisited. *Science Communication*, 6(1), 5–36.
- Schön, D. A., & Rein, M. (1995). *Frame Reflection: Toward the Resolution of Intractable Policy Controversies*. Basic Books.
- Schwendinger, H., & Schwendinger, J. R. (1974). *The sociologists of the chair: a radical analysis of the formative years of North American sociology (1883-1922)*. Basic Books.
- Scott, J. (2008). Modes of power and the re-conceptualization of elites. *The Sociological Review*, 56, 25–43. <https://doi.org/10.1111/j.1467-954X.2008.00760.x>
- Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.
- Sechrest, L., & Sidani, S. (1995). Quantitative and qualitative methods: Is There an Alternative? *Evaluation and Program Planning*, 18(1), 77–87. [https://doi.org/10.1016/0149-7189\(94\)00051-X](https://doi.org/10.1016/0149-7189(94)00051-X)
- Selinger, E., & Crease, R. P. (2006). *The Philosophy of Expertise*. New York: Columbia University Press.
- Sen, A. (1992). *The political economy of targeting*. World Bank Washington, DC. Retrieved from [https://www.researchgate.net/profile/Hindriks\\_Jean/publication/5155655\\_The\\_Political\\_Economy\\_of\\_Targeting/links/53f4a0d50cf22be01c3ed4ea.pdf](https://www.researchgate.net/profile/Hindriks_Jean/publication/5155655_The_Political_Economy_of_Targeting/links/53f4a0d50cf22be01c3ed4ea.pdf)
- Shapin, S., & Schaffer, S. (2011). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (With a New introduction by the authors edition). Princeton, N.J: Princeton University Press.
- Sikkink, K. (2012). *Ideas and Institutions: Developmentalism in Brazil and Argentina*. Cornell University Press.
- Simon, H. A. (1997). *Administrative Behavior, 4th Edition*. Simon and Schuster.
- Sinclair, T. J. (2008). *The New Masters of Capital: American Bond Rating Agencies and the Politics of Creditworthiness*. Ithaca, NY: Cornell University Press.
- Sismondo, S. (2011). *An Introduction to Science and Technology Studies*. John Wiley & Sons.
- Skocpol, T. (1985). *Bringing the State Back In*. New York and Cambridge: Cambridge University Press.
- Somerville, R. A., & Taffler, R. J. (1995). Banker judgement versus formal forecasting models: The case of country risk assessment. *Journal of Banking & Finance*, 19(2), 281–297. [https://doi.org/10.1016/0378-4266\(94\)00051-4](https://doi.org/10.1016/0378-4266(94)00051-4)

- Stake, R. E. (1983). The Case Study Method in Social Inquiry. In *Evaluation Models* (pp. 279–286). Springer, Dordrecht. Retrieved from [https://link.springer.com/chapter/10.1007/978-94-009-6669-7\\_16](https://link.springer.com/chapter/10.1007/978-94-009-6669-7_16)
- Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387–420. <https://doi.org/10.1177/030631289019003001>
- Steen, M. (2015). Upon Opening the Black Box and Finding It Full: Exploring the Ethics in Design Practices. *Science, Technology, & Human Values*, 40(3), 389–420. <https://doi.org/10.1177/0162243914547645>
- Stehr, N., & Grundmann, R. (2011). *Experts: The Knowledge and Power of Expertise*. Routledge.
- Sternberg, E. (1993). Justifying Public Intervention without Market Externalities: Karl Polanyi's Theory of Planning in Capitalism. *Public Administration Review*, 53(2), 100–109. <https://doi.org/10.2307/976702>
- Sternberg, E. (1996). Recuperating from Market Failure: Planning for Biodiversity and Technological Competitiveness. *Public Administration Review*, 56(1), 21–29. <https://doi.org/10.2307/3110050>
- Stern, N. H., & Ferreira, F. (1993). *The World Bank as "intellectual Actor."* Development Economics Research Programme, Suntory-Toyota International Centre for Economics and Related Disciplines.
- Stiglitz, J. E. (1999). Knowledge as a Global Public Good. In *Global Public Goods* (pp. 308–325). Oxford University Press. Retrieved from <http://www.oxfordscholarship.com/view/10.1093/0195130529.001.0001/acprof-9780195130522-chapter-15>
- Stiglitz, J. E. (2000). *Economics of the Public Sector*. W. W. Norton.
- Swyngedouw, E. (2005). Governance Innovation and the Citizen: The Janus Face of Governance-beyond-the-State. *Urban Studies*, 42(11), 1991–2006. <https://doi.org/10.1080/00420980500279869>
- Swyngedouw, E. (2009). The Antinomies of the Postpolitical City: In Search of a Democratic Politics of Environmental Production. *International Journal of Urban and Regional Research*, 33(3), 601–620. <https://doi.org/10.1111/j.1468-2427.2009.00859.x>
- Swyngedouw, E. (2014). *The Post-Political and Its Discontents: Spaces of Depoliticization, Spectres of Radical Politics*. Edinburgh University Press.
- Tarnoff, C. (2015, March 11). Millennium Challenge Corporation [Report]. Retrieved March 19, 2018, from <https://digital.library.unt.edu/ark:/67531/metadc505542/>

- Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214. [https://doi.org/10.1016/0167-2231\(93\)90009-L](https://doi.org/10.1016/0167-2231(93)90009-L)
- Tendler, P. J. (1977). *Inside Foreign Aid*. Baltimore, Md: The Johns Hopkins University Press.
- The Evaluation Gap Working Group. (2006). *When Will We Ever Learn? Improving Lives Through Impact Evaluation*. Center For Global Development. Retrieved from <https://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation>
- The World Bank. (2015). *Annual Report 2015* (Text/HTML). The World Bank. Retrieved from <http://www.worldbank.org/en/about/annual-report-2015>
- Timmermans, S., & Epstein, S. (2010). A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology*, 36(1), 69–89. <https://doi.org/10.1146/annurev.soc.012809.102629>
- Toulmin, S. E. (2009). *Return to Reason*. Harvard University Press.
- Tugwell, R. G. (1954). *The place of planning in society: Seven lectures on the place of planning in society with special reference to Puerto Rico*. San Juan: Puerto Rico Planning Board. Retrieved from <https://catalog.hathitrust.org/Record/101045970>
- Turner, S. (2001). What is the Problem with Experts? *Social Studies of Science*, 31(1), 123–149. <https://doi.org/10.1177/030631201031001007>
- Turner, S. (2003). *Liberal Democracy 3.0: Civil Society in an Age of Experts*. SAGE.
- Turner, S. (2007). Political epistemology, experts, and the aggregation of knowledge. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 1(1), 36.
- Turner, S. (2008). Balancing expert power: Two models for the future of politics. *Knowledge and Democracy: A 21st-Century Perspective*, 119–141.
- Turner, S. (2013). *The Politics of Expertise*. New York: Routledge.
- United Nations. (2000). Millenium Declaration. Retrieved July 8, 2017, from <http://www.un.org/en/development/devagenda/millennium.shtml>
- University of Maryland. (1999). National Security Council Project. Retrieved April 23, 2018, from <http://www.cissm.umd.edu/projects/national-security-council-project>
- Uphoff, N. (1989). Distinguishing Power, Authority & Legitimacy: Taking Max Weber at His Word by Using Resources-Exchange Analysis. *Polity*, 22(2), 295–322. <https://doi.org/10.2307/3234836>

- Valente, T. W., & Rogers, E. M. (1995). The Origins and Development of the Diffusion of Innovations Paradigm as an Example of Scientific Growth. *Science Communication*, 16(3), 242–273. <https://doi.org/10.1177/1075547095016003002>
- Van De Walle, N. (1999). Aid's Crisis of Legitimacy: Current Proposals and Future Prospects. *African Affairs*, 98(392), 337–352.
- Vasquez, I. (1996). Brady Plan and Market-Based Solutions to Debt Crises, The. *Cato J.*, 16, 233.
- Viscusi, W. K. (2009). The devaluation of life. *Regulation & Governance*, 3(2), 103–127. <https://doi.org/10.1111/j.1748-5991.2009.01052.x>
- Wallerstein, I. (Ed.). (1996). *Open the Social Sciences: Report of the Gulbenkian Commission on the Restructuring of the Social Sciences* (1 edition). Stanford, Calif: Stanford University Press.
- Watson, S. (1991). Femocratic feminisms. *The Sociological Review*, 39(S1), 186–204. <https://doi.org/10.1111/j.1467-954X.1991.tb03360.x>
- Watson, V. (2003). Conflicting rationalities: implications for planning theory and ethics. *Planning Theory & Practice*, 4(4), 395–407. <https://doi.org/10.1080/1464935032000146318>
- Watts, M. J. (1993). Development I: power, knowledge, discursive practice. *Progress in Human Geography*, 17(2), 257–272. <https://doi.org/10.1177/030913259301700210>
- Weber, M. (2009). *The Theory Of Social And Economic Organization*. Simon and Schuster.
- Weiss, C. H. (1979). The Many Meanings of Research Utilization. *Public Administration Review*, 39(5), 426–431. <https://doi.org/10.2307/3109916>
- Wetherell, M., Taylor, S. J. A., & Yates, S. J. (Eds.). (2001). *Discourse Theory and Practice: A Reader* (1 edition). London ; Thousand Oaks, Calif.: SAGE Publications Ltd.
- Whitfield, L. (2009). *The Politics of Aid: African Strategies for Dealing with Donors*. Oxford University Press.
- Wildavsky, A. B. (1979). *Speaking truth to power: the art and craft of policy analysis*. Little, Brown.
- Williams, A., & Siddique, A. (2008). The use (and abuse) of governance indicators in economics: a review. *Economics of Governance*, 9(2), 131–175. <https://doi.org/10.1007/s10101-006-0025-9>
- Williamson, J. (1965). Regional Inequality and the Process of National Development: A Description of the Patterns. *Economic Development and Cultural Change*, 13(4, Part 2), 1–84. <https://doi.org/10.1086/450136>



- Williamson, J. (2003). The Washington consensus and beyond. *Economic and Political Weekly*, 1475–1481.
- Williamson, J. (2004). The Strange History of the Washington Consensus. *Journal of Post Keynesian Economics*, 27(2), 195–206. <https://doi.org/10.2307/4538920>
- Willis, C. (1995). *Form Follows Finance: Skyscrapers and Skylines in New York and Chicago*. Princeton Architectural Press.
- Wilson, G. (2006). Beyond the technocrat? The professional expert in development practice. *Development and Change*, 37(3), 501–523.
- World Bank (Ed.). (1998). *Assessing aid: what works, what doesn't, and why*. Washington, D.C: Oxford University Press.
- Yapa, L. (1996). What Causes Poverty?: A Postmodern View. *Annals of the Association of American Geographers*, 86(4), 707–728. <https://doi.org/10.1111/j.1467-8306.1996.tb01773.x>
- Yin, R. K. (1984). *Case study research: design and methods*. Sage Publications.
- Yonay, Y. (1994). When Black Boxes Clash: Competing Ideas of What Science Is in Economics, 1924-39. *Social Studies of Science*, 24(1), 39–80. <https://doi.org/10.1177/030631279402400103>
- Yonay, Y. (1998). *The Struggle over the Soul of Economics* (First Edition). Princeton, N.J: Princeton University Press.
- Zald, M. N., & Lounsbury, M. (2010). The Wizards of Oz: Towards an Institutional Approach to Elites, Expertise and Command Posts. *Organization Studies*, 31(7), 963–996. <https://doi.org/10.1177/0170840610373201>
- Ziliak, S., & Teather-Posadas, E. R. (2014). The unprincipled randomization principle in economics and medicine. *Ziliak, Stephen T., and Teather-Posadas, Edward R.* Retrieved from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2396806](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2396806)
- Žižek, S. (2014). *The Universal Exception*. Bloomsbury Publishing.

## Government Documents and Other Primary Source Material

- 3ie (2015). Impact Evaluation Repository. Available:  
<http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/>.
- Adventist Development and Relief Association (2011). *Commercial Development of Farmer-Based Organizations (CDFO)*. Presentation at M&E Technical Dissemination Workshop at the Novotel Hotel, Accra, Ghana on February 10, 2011.
- Anderson, S., Breitbarth, T., & Osborne, T. (2010). *MCC's Experience with the Analysis of Constraints to Growth: Survey Results*. Washington, DC: Millennium Challenge Corporation.
- Annan, K. (2002, March 21). *Aid Does Work*. Statement by Secretary-General Kofi Annan to the International Conference on Financing for Development, Monterrey, Mexico. Retrieved from <http://www.un.org/ffd/pressrel/21a.htm>.
- Asian Development Bank (2004). *Review of the Asian Development Bank's Policy on the Performance-Based Allocation of Asian Development Fund Resources*. Manila: Asian Development Bank.
- Asian Development Bank (2008). *Refining the Performance-Based Allocation of Asian Development Fund Resources*. Manila: Asian Development Bank.
- Asian Development Bank (2010). *ADF's Performance-Based Allocation System: Review of the Current System and the Post-Conflict Assistance Phaseout*. Manila: Asian Development Bank.
- Beatty, A.J. (2006, April 27). *How does the Ghana impact evaluation fit into overall M&E activities? Why do we care about randomization?* Electronic communication sent to the Ghana Transaction Team. Washington, DC: Millennium Challenge Corporation.
- Boucher, S., Malacarne, J. & Carter, M. (2016, June 30). *Understanding Impact Heterogeneity in Nicaragua's Rural Business Development Program*. Presentation delivered at Deep Dive into MCC's Nicaragua Farmer Training Impact Evaluation. Washington, DC: Millennium Challenge Corporation.
- Bush, G. W. (2002a, March 14). *Remarks to the Inter-American Development Bank*. Speech presented at The Inter-American Development Bank, Washington, DC. Retrieved from <https://2001-2009.state.gov/e/eeb/rls/rm/2002/8780.htm>.
- Campuzano, L. & Blair, R. (2011). *Final Impact Evaluation Design for the Production and Business Services Activity of the Productive Development Project – Revised* (ESVED-264). Princeton, NJ: Mathematica Policy Research.
- Carter, M. (2016, June 30). *Understanding Impact Heterogeneity in Nicaragua's Rural Business*

- Development Program*. Remarks delivered at Deep Dive into MCC's Nicaragua Farmer Training Impact Evaluation. Washington, DC: Millennium Challenge Corporation.
- Center for Global Development (n.d.). *Groups and Individuals Consulted*. Closing the Evaluation Gap. Washington, DC: Center for Global Development.
- Center for Global Development (2002). *Draft invite for senior-level workshops on the Millennium Challenge Account*. Washington, DC: Center for Global Development.
- Combating Multilateral Development Bank Corruption: U.S. Treasury Role and Internal Efforts: Hearing Before the Committee on Foreign Relations (S. Hrg. 108-734), U.S. Senate, 108<sup>th</sup> Cong. 7 (2004) (Testimony of John B. Taylor, Under Secretary for International Affairs).
- Council of Economic Advisors (2015). *Kroszner, Randall (Randy) – Subject Files*. Trade: Millennium Challenge Account [Folder]. Dallas, TX: George W. Bush Presidential Library and Museum.
- Dam, K. W. (2002, June 22). *Treasury as a Foreign Policy Institution*. Speech presented at The Council on Foreign Relations, Chicago, IL. Retrieved from <https://www.treasury.gov/press-center/press-releases/Pages/po3220.aspx>
- Dennis, B. (2006). *Proposed General Principles for Initial Economic Analysis of Proposed Projects*. Washington, DC: Millennium Challenge Corporation.
- Farley, K., Lucas, S., Molyneaux, J., Penn, K. & Hogue, E. (2012). *Principles into Practice – Impact Evaluations of Agriculture Projects*. Washington, DC: Millennium Challenge Corporation.
- Feld, S. (2009). *Constraints Analysis Review*. Washington, DC: Millennium Challenge Corporation.
- Foreign Affairs Budget: Hearing before the Committee on Foreign Relations (S. Hrg. 108-19), U.S. Senate, 108<sup>th</sup> Cong. (2003).
- Fortson, K. & Rangarajan, A. (2008). *Water-to-Market Evaluation Plan (MCC-008)*. Princeton, NJ: Mathematica Policy Research.
- Hubbard, G. R. (2002, July 12). *The Millennium Challenge Account: Taking Governance and Growth Seriously*. Remarks at Making Sustainable Development Work: Governance, Finance, and Public-Private Cooperation. Washington, DC: US Department of State. Retrieved from <https://www0.gsb.columbia.edu/faculty/ghubbard/speeches/7.12.02.pdf>
- ISSER (n.d.). *Approach to Evaluating Impact of the training component of the MiDA Programme*. Accra: Institute of Statistical, Social and Economic Research.
- ISSER (2007). *Proposal for GLSS5+ and FBO Surveys and Impact Evaluation of the MiDA Program*. Accra: Institute of Statistical, Social and Economic Research.

- ISSER (2012). *An Impact Evaluation of the MiDA FBO Training: Final Report*. Accra: Institute of Statistical, Social and Economic Research.
- J-PAL (2006). *Evaluating Social Programs*. Course Binder for Executive Training at J-PAL May 30–June 3, 2006. Cambridge, MA: Massachusetts Institute of Technology.
- Jacquet, P. (2006). Evaluations and Aid Effectiveness. *Rescuing the World Bank: a CGD working group report and selected essays*, 171–182. Washington, DC: Center for Global Development.
- Kelly, T. (2007). *Integrated Rural Development*. Development Policy Discussion Paper. Washington, DC: Millennium Challenge Corporation.
- Leonard, K. (2011, July 21). *Agricultural Development Program: Lessons Learned from the Project Evaluation in Burkina Faso*. Presentation at the Conference on Agriculture Impact Evaluations: Learning What Works. Washington, DC: Millennium Challenge Corporation.
- Levine, D. I. (2006). *Learning What Works – and What Doesn't: Building Learning into the Global Aid Industry*. Unpublished manuscript. Berkeley, CA: University of California at Berkeley. Retrieved from <http://faculty.haas.berkeley.edu/levine/>.
- MCA-Benin. (2012b). *4: Test analyses for Roads*. Cotonou: Millennium Challenge Account-Benin.
- MCA-Benin (2012a). *Corruption – Test 2*. Cotonou: Millennium Challenge Account-Benin.
- MCC (n.d.a) *CA Consultations – Lessons from other MCC countries*. Washington, DC: Millennium Challenge Corporation.
- MCC (n.d.b) *Consultative Process in El Salvador*. Washington, DC: Millennium Challenge Corporation.
- MCC (n.d.c). *Introduction to Constraints Analysis (a.k.a. Growth Diagnostics)*. Washington, DC: Millennium Challenge Corporation.
- MCC (2005a). *Summary of the Millennium Challenge Compact – Honduras*. Washington, DC: Millennium Challenge Corporation.
- MCC (2005b). *Working with MCC: Overview of Steps and Processes*. Washington, DC: Millennium Challenge Corporation.
- MCC (2006a). *Ghana Compact - Impact Evaluation Design Session led by Professors Ernest Aryeetey and Christopher Udry: Proposed Agenda*. Washington, DC: Millennium Challenge Corporation.

- MCC (2006b). *Guidelines for Monitoring and Evaluation Plans in Fiscal Year 2007 Guidelines for Compact Eligible Countries*. Washington, DC: Millennium Challenge Corporation.
- MCC (2006c, December 21). *Request for Proposal. Request MCC-07-RFP-0008*. Washington, DC: Millennium Challenge Corporation.
- MCC (2007a). *Aide Memoire: El Salvador Implementation Oversight. M&E Planning and Impact Evaluation Mission. June 21-30, 2007*. Washington, DC: Millennium Challenge Corporation.
- MCC (2007b). *Trip Report for M&E Planning Mission, June 21 to 30, 2007*. Washington, DC: Millennium Challenge Corporation.
- MCC (2009). *Guidelines for Conducting a Constraints Analysis*. Washington, DC: Millennium Challenge Corporation.
- MCC (2011). *Agriculture Impact Evaluations: Learning What Works – Agenda*. Washington, DC: Millennium Challenge Corporation.
- MCC (2012a). *MCC Impact Evaluation Learning Event – Agriculture Evaluations Portfolio Peer Review Workshop: Agenda*. Washington, DC: Millennium Challenge Corporation.
- MCC (2012b). *MCC’s First Impact Evaluations: Farmer Training Activities in Five Countries*. Washington, DC: Millennium Challenge Corporation.
- MCC (2012c). *Measuring Results of the El Salvador Production and Business Services Investments*. Washington, DC: Millennium Challenge Corporation.
- MCC (2013). *Guidelines for Conducting a Constraints Analysis*. Millennium Challenge Corporation.
- MCC and USAID (2011). *Guidelines: Impact Evaluations of Agriculture Projects*. Washington, DC: Millennium Challenge Corporation.
- MCC Department of Policy & Evaluation (2010). *Analyzing Constraints to Growth in Selected Countries*. Washington, DC: Millennium Challenge Corporation.
- MCC Development Policy (2005). *Action Memo: MCA Country Selection Criteria and Methodology for FY 2006*. Washington, DC: Millennium Challenge Corporation.
- Ministry of Finance and National Planning (2010). *An Analysis of Constraints to Inclusive Growth in Zambia – Final Draft*. Lusaka: Ministry of Finance and National Planning.
- Multilateral development banks: promoting effectiveness and fighting corruption: Hearing before the Committee on Foreign Relations (S. Hrg. 109-913), U. S. Senate (2006).
- National Millennium Compact Development Project (2013). *Liberia Constraints Analysis*. Monrovia: Government of Liberia.

- New African Initiative (2001). *A New African Initiative – Merger of the Millennium Partnership for the African Recovery Programme (MAP) and Omega Plan*. Pretoria.
- NORC (2011). *Draft Final Report: Impact Evaluation of the Farmer Training and Development Activity in Honduras*. Chicago, IL: National Opinion Research Center.
- NORC (2013). *Final Report (Revised): Impact Evaluation of the Farmer Training and Development Activity in Honduras*. Chicago, IL: National Opinion Research Center.
- Partnership for Growth (2011). *Tanzania Growth Diagnostic: A Joint Analysis for the Governments of the United Republic of Tanzania and the United States of America*. Washington, DC: Millennium Challenge Corporation.
- Performance Indicator Working Group (2002a). *Draft Work Plan – April 5, 2002*. Washington, DC: US Department of the Treasury.
- Performance Indicator Working Group (2002b). *Follow-up on Steering Group Requests*. Washington, DC: US Department of the Treasury.
- Poverty Action Lab (2005). *Evaluating Social Programs*. Course Binder for Executive Training at the Poverty Action Lab. March 21–25, 2005. Cambridge, MA: Massachusetts Institute of Technology.
- Replenishment authorizations for the World Bank’s International Development Association, the Asian Development Fund, and the African Development Fund: Hearing before the Subcommittee on International Economic Policy, Export and Trade Promotion of the Committee on Foreign Relations (S. Hrg. 107-796), U.S. Senate 107<sup>th</sup> Cong. 5 (2002) (Testimony of John B. Taylor, Under Secretary of the Treasury for International Affairs).
- Rigorous Evidence: Key to Progress Against World Poverty?* Millennium Challenge Corporation. October 29, 2008. [Transcript]. Retrieved from <https://assets.mcc.gov/content/uploads/2017/05/transcript-102908-poverty.pdf>
- Rubin, R. (1993). *Memorandum for the President: Year-End Thoughts on the NEC*. Washington, DC: The White House.
- Seidman, E. (1996). *Memorandum for Tom O’Donnell: The NEC*. Washington, DC: The White House.
- Seitz, V. & Adato, M. (2004). *Gender Equality and Poverty Reduction through Growth*. Washington, DC: Millennium Challenge Corporation.
- Taylor, J. B. (2002a, October 2). *Finding New Business Models at the Multilateral Development Banks*. Speech presented at the Business Council for International Understanding, Washington, DC. Retrieved from <https://web.stanford.edu/~johntayl/Chronological%20List.htm>.

- Taylor, J. B. (2002b, February 19). *Improving the Regional Development Banks*. Conference on Financing For Development on Regional Challenge and The Regional Development Banks at the Peterson Institute for International Economics, Washington, DC. Retrieved from <https://web.stanford.edu/~johntayl/Chronological%20List.htm>
- Taylor, J. B. (2002c, April 30). *New Policies for Economic Development*. Presentation given at the Annual Bank Conference on Development Economics, Washington, DC. Retrieved from <https://web.stanford.edu/~johntayl/Chronological%20List.htm>
- Taylor, J. B. (2004a, October 26). *Getting International Economic Development Right: Is Effective Foreign Assistance Possible?* Speech presented at the Cleveland Council on World Affairs at Case Western Reserve University, Cleveland, OH. Retrieved from <https://web.stanford.edu/~johntayl/Chronological%20List.htm>
- Taylor, J. B. (2005, January 28). *The Private Sector's Role in Promoting Economic Growth in the Broader Middle East and North Africa*. Speech presented at the World Economic Forum Business Dialogue: The Arab Business Council and the G-8/BMENA Forum for the Future, Davos, Switzerland. Retrieved from <https://web.stanford.edu/~johntayl/Chronological%20List.htm>
- The Millennium Challenge Account: a new way to aid: Hearing before the Committee on Foreign Relations (S. Hrg. 108-25), U.S. Senate, 108<sup>th</sup> Cong. (2003).
- The Millennium Challenge Account: Hearing before the Committee on International Relations (Serial no. 108-12), U.S. House of Representatives (2003).
- Treasury (n.d.). *MCA Country Selection Methodology*. Washington, DC: US Department of the Treasury
- Treasury (2002a). *MCA Inter-Agency Meeting – Governance Sub-Group Discussion Items*. Washington, DC: US Department of the Treasury.
- Treasury (2002b). *Treasury Governance Indicator*. Washington, DC: US Department of the Treasury.
- United Nations (2000, September 28). *United Nations Millennium Declaration*. Resolution adopted by the General Assembly at its Fifty-fifth session (55/2), New York, NY. Retrieved from <http://www.un.org/millennium/declaration/ares552e.htm>
- United Nations Development Programme (2004). *Governance Indicators a Users' Guide*. Oslo: UNDP Oslo Governance Centre.
- U.S. Congress, 108<sup>th</sup> Congress, Pub. L. 108-199, SEC. 601, Millennium Challenge Act of 2003. Washington, Government Printing Office, 2004.
- Warner, A. (2004). *Cape Verde Growth Report*. Washington, DC: Millennium Challenge Corporation.

Weiner, D. & Lucas, S. (2002, June 5). *Millennium Challenge briefing invitation*. Electronic communication sent to Randall S. Krozsner. Washington, DC: Center for Global Development.

*'When Will We Ever Learn: Improving Lives through Impact Evaluation' Policy Recommendations from the CGD Evaluation Gap Working Group*. Center for Global Development. May 31, 2006 [Transcript]. Retrieved from <https://www.cgdev.org/doc/events/5.31.06/5.31.06EvalGapTrans.pdf>

White House (2001). *National Security Presidential Directive 1: Organization of the National Security Council System*. Washington, DC: The White House.

White House (2002a, March 14). *Background Briefing by Senior Administration Officials on the President's Global Development Aid Policy*. Washington, DC: The White House Office of the Press Secretary.

White House (2003). The Millennium Challenge Account: Fact Sheet, *Economic Perspectives* 8(2). Washington, DC: US Department of State.

Wiebe, F. (2007, March 19). *Economic Analysis in the MCC*. Presentation delivered at MCC-University, Washington, DC.

World Bank (2001). *World Bank Lending Instruments: Resources for Development Impact*. Washington, DC: The World Bank Group Operations Policy and Country Services



## Appendix A: List of Interviewees

### Case 1: Governance Indicators

- 1A. Anonymous. Deputy Assistant Secretary, Department of State. June 1, 2015.
- 1B. Anonymous. Director, National Security Council. June 11, 2015.
- 1C. Anonymous. Director, Department of State. June 16, 2015.
- 1D. Anonymous. Deputy Assistant Secretary, Department of State. June 22, 2015.
- 1E. Anonymous. Deputy Assistant Secretary, Department of the Treasury. June 28, 2015.
- 1F. Anonymous. Director, National Security Council. July 2, 2015.
- 1G. Anonymous. Senior Staffer, United States Senate Foreign Relations Committee.  
July 14, 2015.
- 1H. Anonymous. Director, Department of the Treasury. July 19, 2015.
- 1I. Anonymous. Private Sector Development Specialist, International Finance Corporation.  
July 22, 2015.
- 1J. Anonymous. Director, Department of the Treasury. August 5, 2015.
- 1K. Anonymous. Undersecretary, Department of State. August 11, 2015.
- 1L. Anonymous. Economist, World Bank Development Economics Research Group.  
August 12, 2015.
- 1M. Anonymous. Administrator, United States Agency for International Development.  
August 28, 2015.
- 1N. Anonymous. Director, United States Agency for International Development.  
September 1, 2015.
- 1O. Anonymous. Undersecretary, Department of the Treasury. September 2, 2015.

## **Case 2: Growth Diagnostics**

- 2A. Anonymous. Lead Economist, Millennium Challenge Account. May 2, 2015.
- 2B. Anonymous. Deputy Vice President, Compact Operations, MCC. February 11, 2016.
- 2C. Anonymous. Director, Economic Analysis. February 18, 2016.
- 2D. Anonymous. Director, Health and Education, MCC. February 19, 2016.
- 2E. Anonymous. Director, Legal and Regulatory Reform, MCC. February 24, 2016.
- 2F. Anonymous. Director, Economic Analysis, MCC. February 23, 2016.
- 2G. Anonymous. Senior Director, Compact Operations, MCC. February 26, 2016.
- 2H. Anonymous. Senior Director, Environment and Social Protection, MCC. March 2, 2016.
- 2I. Anonymous. Lead Economist, MCC. March 4, 2016.
- 2J. Anonymous. Vice President, MCC. March 15, 2016.
- 2K. Anonymous. Director, Compact Development, MCC. March 22, 2016.
- 2L. Anonymous. Program Officer, Agriculture and Land, MCC. March 22, 2016.
- 2M. Anonymous. Economist, MCC. March 22, 2016.
- 2N. Anonymous. Economist, MCC. March 22, 2016.
- 2O. Anonymous. Director, Agriculture and Land, MCC. March 25, 2016.
- 2P. Anonymous. Deputy Vice President, Compact Operations, MCC. March 29, 2016.
- 2Q. Anonymous. Managing Director, Agriculture. March 29, 2016.
- 2R. Anonymous. Director, Compact Development, MCC. March 30, 2016.
- 2S. Anonymous. Director, Infrastructure, MCC. April 4, 2016.
- 2T. Anonymous. Senior Director, Infrastructure, MCC. April 5, 2016.

### **Case 3: Randomized Controlled Trials**

- 3A. Anonymous. Senior Director, Monitoring and Evaluation, MCC. February 17, 2016.
- 3B. Anonymous. Managing Director, Department of Accountability, MCC. April 27, 2016.
- 3C. Anonymous. Associate Director, Policy and International Relations, MCC. May 26, 2016.
- 3D. Anonymous. Senior Agricultural Specialist, MCC. June 22, 2016.
- 3E. Anonymous. Director, Monitoring and Evaluation, MCC. June 24, 2016.
- 3F. Anonymous. Director, Economic Analysis, MCC. June 27, 2016.
- 3G. Anonymous. Associate Director, Monitoring and Evaluation, MCC. June 29, 2016.
- 3H. Anonymous. Program Officer, Monitoring and Evaluation, MCC. June 29, 2016.
- 3I. Anonymous. Senior Director, Agriculture and Land, MCC. June 30, 2016.
- 3J. Anonymous. Director, Agriculture and Land, MCC. July 6, 2016.
- 3K. Anonymous. Director, Economic Analysis, MCC. July 7, 2016.
- 3L. Anonymous. Associate Director, Agriculture and Land, MCC. July 8, 2016.
- 3M. Anonymous. Project Manager, Development Alternatives Incorporated. July 27, 2016.
- 3N. Anonymous. Project Manager, Adventist Development and Relief Association.  
August 1, 2016.
- 3O. Anonymous. Evaluation Specialist, Institute of Statistical, Social, and Economic Research.  
August 2, 2016.

## Appendix B: Tables and Figures for Governance Indicators Chapter

Table B-1

### *Differences in Income Between Low-income Countries Based on MCA Qualification Criteria Performance, 2004–2012*

Year	Total Candidate Countries		Statutorily Prohibited	"Passed" WGI	"Failed" WGI	"Passed" WGI (%)	Difference (%)	Qualified	Failed to Qualify	Qualified (%)	Difference (%)
	n	GNI / capita (\$)									
2006	32		3	23	9	71.9		8	24	25.0	
		GNI / capita (\$)		2,374	2,117		11.4	2,465	2,247		9.3
2007	34		4	24	10	70.6		8	26	23.5	
		GNI / capita (\$)		2,572	2,794		-8.3	2,778	2,594		6.9
2008	33		5	25	8	75.8		4	29	12.1	
		GNI / capita (\$)		2,663	2,644		0.7	2,660	2,658		0.1
2009	31		2	24	7	77.4		6	25	19.4	
		GNI / capita (\$)		2,862	2,900		-1.3	3,290	2,770		17.2
2010	35		4	26	9	74.3		6	29	17.1	
		GNI / capita (\$)		2,866	2,901		-1.2	2,965	2,856		3.7
2011	32		3	26	6	81.3		2	30	6.3	
		GNI / capita (\$)		2,902	3,017		-3.9	2,865	2,927		-2.1
2012	30		1	22	8	73.3		7	23	23.3	
		GNI / capita (\$)		2,797	3,073		-9.4	2,873	2,869		0.1
Average Sample Sizes (2006 - 2012)	32		3	24	8	74.9		6	27	18.1	
Average Income Levels (2006 - 2012)		GNI / capita (\$)		2,719	2,778		-1.7	2,842	2,703		5.0

*Source:* Performance data is from the Millennium Challenge Corporation. Income data is from the World Bank's World Development Indicators.

*Notes:* Income data at the time of selection was lagged by two years, i.e. selection for fiscal year 2008 used 2006 GNI/capita figures. Data for North Korea was not available in any year. Somalia's income is in GDP per capita.

Table B-2

*Differences in Income Between Lower Middle-income Countries Based on MCA Qualification Criteria Performance, 2006–2012*

Year	Total Candidate Countries		Statutorily Prohibited	"Passed" WGI	"Failed" WGI	"Passed" WGI (%)	Difference (%)	Qualified	Failed to Qualify	Qualified (%)	Difference (%)
	n	72									
2004	GNI / capita (\$)		9	52	20	72.2	47.0	19	53	26.4	42.3
	n		9	599	371			650	423		
2005	GNI / capita (\$)		14	60	22	73.2	31.5	25	57	30.5	57.9
	n		14	581	423			780	430		
2006	GNI / capita (\$)		12	55	26	67.9	35.4	26	55	32.1	32.9
	n		12	668	467			746	535		
2007	GNI / capita (\$)		10	55	24	69.6	25.4	21	58	26.6	40.6
	n		10	723	560			894	592		
2008	GNI / capita (\$)		8	51	24	68.0	43.8	21	54	28.0	43.0
	n		8	766	491			912	589		
2009	GNI / capita (\$)		9	53	20	72.6	41.5	19	54	26.0	39.7
	n		9	830	545			999	668		
2010	GNI / capita (\$)		7	44	19	69.8	40.3	15	48	23.8	35.4
	n		7	835	555			974	681		
2011	GNI / capita (\$)		8	43	20	68.3	18.8	15	48	23.8	23.2
	n		8	868	719			976	773		
2012	GNI / capita (\$)		8	44	16	73.3	25.8	14	46	23.3	25.4
	n		8	874	674			994	770		
Average Sample Sizes (2004-12)			9	51	21	70.5	33.6	19	53	27.0	36.8
Average Income Levels (2000-12)			9	749	534			881	607		

*Source:* Performance data is from the Millennium Challenge Corporation. Income data is from the World Bank's World Development Indicators.

*Notes:* Income data at the time of selection was lagged by two years, i.e. selection for fiscal year 2008 used 2006 GNI per capita figures.

Appendix C: Tables and Figures for Growth Diagnostics Chapter

Table C-1

*Distribution of Categorized Data Sources in MCC Constraints Analyses*

<b>Data Source</b>	<b>N</b>	<b>Percent of Total</b>
World Bank World Development Indicators	202	13.6
Central or National Bank	125	8.4
World Bank Enterprise Surveys	113	7.6
International Monetary Fund	92	6.2
World Bank Doing Business Survey	86	5.8
National Statistics Office	75	5.0
Other World Bank	73	4.9
Other National Ministry	70	4.7
World Economic Forum	57	3.8
National Ministry of Finance or Economy	47	3.2
Other United Nations	41	2.8
Investment Climate Survey Diagnostic	41	2.8
Economist Intelligence Unit	38	2.6
Regional Multilateral Development Bank	35	2.4
National Infrastructure Ministry	35	2.4
Unspecified	33	2.2
National Labor Survey	30	2.0
UN Center for Trade and Development	30	2.0
National Economic Social Survey	28	1.9
Other International Organization	25	1.7
US Based University or Think Tank	25	1.7
Individual Scholars	22	1.5
National Regulatory Agency	20	1.3
United States Government	20	1.3
Other National Survey	19	1.3
UN Food and Agriculture Organization	19	1.3
World Bank Worldwide Governance Indicators	15	1.0
UN World Health Organization	14	0.9
US Based Consultancy	14	0.9

Other International Survey	13	0.9
Other	12	0.8
National Utility or Port Authority	10	0.7
Domestic University or Think Tank	8	0.5
Total*	1,487	100.0

\* The number of data sources is higher than analyses because some analyses use more than one data source.

*Source:* MCC Constraints Analyses' Final Reports.

Table C-2

*Distribution of Categorized Data Sources in MCC Constraints Analyses for High Cost of Finance Constraints*

Data Source	Constraint		
	Bad International Finance (%) (N=41)	Low Domestic Savings (%) (N=87)	Poor Intermediation (%) (N=122)
IMF and National Central Banks	22	52	42
Worldwide Development Indicators	22	25	9
World & Regional Development Banks	10	6	2
Competitiveness Surveys	0	1	14
UN or Other International Organization	12	0	1
Economist Intelligence Unit	7	9	5
National Ministry or Regulator	7	0	10
National Survey	7	5	8
US Government, Think Tank, or Consultancy	7	1	2
International Survey	0	0	1
Other	2	0	0
Unspecified	2	1	6
<b>Total</b>	100	100	100

*Source:* MCC Constraints Analyses' Final Reports.

*Notes:* IMF (International Monetary Fund), WDI (World Bank World Development Indicators), WGI (World Bank Worldwide Governance Indicators), WB (World Bank), EIU (Economist Intelligence Unit), USG (United States Government).



Table C-3

*Distribution of Categorized Data Sources in MCC Constraints Analyses for Market Failure Constraints*

Data Source	Constraint	
	Information Externalities (%) (N=96)	Coordination Externalities (%) (N=11)
Competitiveness Surveys	27	0
National Survey	6	36
UN or Other International Organization	27	18
Other World & Regional Development Bank	7	18
US Government, Think Tank, or Consultancy	3	18
National Ministry or Regulator	10	9
Individual Scholars	8	0
Economist Intelligence Unit	1	0
Worldwide Development Indicators	4	0
IMF and National Central Banks	2	0
International Survey	2	0
Unspecified	1	0
Total	100	100

*Source:* MCC Constraints Analyses' Final Reports.

*Notes:* IMF (International Monetary Fund), WDI (World Bank World Development Indicators), WGI (World Bank Worldwide Governance Indicators), WB (World Bank), EIU (Economist Intelligence Unit), USG (United States Government). *Source:* MCC Constraints Analyses' Final Reports.

Table C-4

*Distribution of Categorized Data Sources in MCC Constraints Analyses for Government Failure Constraints*

Data Source	Constraint	
	Micro Risks (%) (N=222)	Macro Risks (%) (N=145)
Competitiveness Surveys	47	3
IMF and National Central Banks	4	41
National Ministry or Regulator	11	12
Economist Intelligence Unit	0	10
Worldwide Development Indicators	6	9
Worldwide Governance Indicators	7	0
Other World and Regional Development Banks	3	8
National Survey	3	6
Individual Scholars	3	0
UN or Other International Organization	5	3
International Survey	3	0
US Government, Think Tank, or Consultancy	4	1
Unspecified	3	2
Other	0	2
Domestic Think Tank or Consultancy	1	1
Total	100	100

*Source:* MCC Constraints Analyses' Final Reports.

*Notes:* IMF (International Monetary Fund), WDI (World Bank World Development Indicators), WGI (World Bank Worldwide Governance Indicators), WB (World Bank), EIU (Economist Intelligence Unit), USG (United States Government).

Table C-5

*Distribution of Categorized Data Sources in MCC Constraints Analyses for Low Social Return Constraints*

Data Source	Constraint		
	Poor Geography (%) (N=84)	Low Human Capital (%) (N=179)	Bad Infrastructure (%) (N=342)
Competitiveness Surveys	1	20	28
IMF and National Central Banks	1	1	1
National Ministry or Regulator	19	7	21
National Survey	4	31	4
Individual Scholars	0	4	0
International Survey	2	1	0
Other	5	0	1
Other World & Regional Development Banks	12	3	13
Domestic Think Tank or Consultancy	0	1	1
UN or Other International Organization	13	13	9
Unspecified	2	1	2
US Government, Think Tank, or Consultancy	18	3	4
Worldwide Development Indicators	23	15	16
Total	100	100	100

*Source:* MCC Constraints Analyses' Final Reports.

*Notes:* IMF (International Monetary Fund), WDI (World Bank World Development Indicators), WGI (World Bank Worldwide Governance Indicators), WB (World Bank), EIU (Economist Intelligence Unit), USG (United States Government).

Table C-6

*Summary Statistics on Analysis Types Found in MCC Constraints Analyses*

	Snapshot		Time Series		Total	
	N	%	N	%	N	%
Descriptive	190	14.0	429	31.7	619	45.7
Benchmarked	397	29.3	239	17.7	636	47.0
Correlation	60	4.4	21	1.6	81	6.0
Ordinary Least Squares	7	0.5	4	0.3	11	0.8
Map	7	0.5	0	0.0	7	0.5
Total	661	48.8	693	51.2	1,354	100.0

*Source:* MCC Constraints Analyses' Final Reports.

## Appendix D: Tables and Figures for Randomized Controlled Trials Chapter

Table D-1

*List of Completed and Ongoing RCTs at MCC as of Fiscal Year 2014–2015*

Country	Project / Activity Type	Sector	Intervention Type	Methodology	Status
Armenia	Agricultural Training and Input Subsidy	ARD	Education & Agricultural input subsidy	RRO	Complete
Benin	Access to Land	Land	Village-based land formalization schemes	RA	Complete
Burkina Faso	Water Management and Irrigation	ARD	Individual land parcels with irrigation & title	RA	Ongoing
El Salvador	Rural Electrification Financing	Energy	Household connection subsidy	RPL	Complete
El Salvador	Education and Training Scholarships	HECS	Individual scholarships	RA	Complete
El Salvador	Agriculture Production and Business Services	ARD	Education & Individual Mentoring	RRO	Complete
El Salvador	Handicrafts Production and Business Services	PSD	Education & Individual Mentoring	RRO	Complete
Georgia	Agribusiness Development	ARD	Business development grants	RA	Converted to quasi-experimental
Ghana	Agricultural Training and Input Subsidy	ARD	Education & Ag input subsidy	RRO	Complete
Honduras	Agriculture Production and Business Services	ARD	Education & Individual Mentoring	RRO	Converted to quasi-experimental
Indonesia	Community Nutrition	HECS	Community based health grants	RA	Ongoing

Lesotho	Rural Water Supply and Sanitation	W&S	Rural borehole provision	RPL	Ongoing
Moldova	Agriculture Production and Business Services	ARD	Education & Individual Mentoring	RA	Converted to quasi-experimental
Mongolia	Peri-Urban Land Leasing	Land	Individual land titles	RA	Ongoing
Mongolia	Vocational Education	HECS	Individual training and access to new equipment	RA	Ongoing
Morocco	Enterprise Support Grants	PSD	Business development grants	RA	Complete
Morocco	Fruit Tree Rehabilitation	ARD	Education & Individual Mentoring	RA	Complete
Namibia	Vocational and Skills Training	HECS	Vocational education	RA	Ongoing
Namibia	Community Based Rangeland & Livestock Management	ARD	Education & Individual Mentoring	RA	Ongoing
Nicaragua	Rural Business Development Services	ARD	Education & Individual Mentoring	RRO	Complete
Nicaragua	Agricultural Technical and Financial Assistance	ARD	Individual agricultural credit	RRO	Complete
Philippines	Community Development Grants	HECS	Community based social development grants	RA	Ongoing
Tanzania	Rural Electrification Financing	Energy	Household connection subsidy	RA	Ongoing

*Source:* MCC Impact Evaluation Database.

*Notes:* Only RCTs where a control group was identified are included in the list. Randomized designs include RRO (Randomized Roll-out), RA (Random assignment), and RPL (Randomized pipeline). Sector acronyms stand for Agricultural and Rural Development (ARD), Private Sector Development (PSD), Health, Education, and Community Services (HECS), and Water and Sanitation (W&S).