

The Seductions of Quantification

**MEASURING HUMAN RIGHTS, GENDER VIOLENCE,
AND SEX TRAFFICKING**

Sally Engle Merry

The University of Chicago Press Chicago and London

Contents

Acknowledgments vii

CHAPTER ONE

A World of Quantification 1

CHAPTER TWO

Indicators as a Technology of Knowledge 27

CHAPTER THREE

Measuring Violence against Women 44

CHAPTER FOUR

Categorizing Violence against Women: The Cultural Work of
Commensuration 75

CHAPTER FIVE

Measuring the Unmeasurable: The *US Trafficking in Persons Reports* 112

CHAPTER SIX

Knowledge Effects and Governance Effects of the *Trafficking in
Persons Reports* 140

CHAPTER SEVEN

Human Rights Indicators: Translating Law into Policy 161

CHAPTER EIGHT

Conclusions 207

Notes 223 References 225 Index 243

A World of Quantification

Quantification is seductive. It offers concrete, numerical information that allows for easy comparison and ranking of countries, schools, job applicants, teachers, and much else. It organizes and simplifies knowledge, facilitating decision making in the absence of more detailed, contextual information. By quantification, I mean the use of numbers to describe social phenomena in countable and commensurable terms. Quantification depends on constructing universal categories that make sense across national, class, religious, and regional lines. Categorized numbers can then be bundled together into more complex representations of social phenomena, such as good governance or the rule of law. These numbers convey an aura of objective truth and scientific authority despite the extensive interpretive work that goes into their construction.

Indeed, it is the capacity of numbers to provide knowledge of a complex and murky world that renders quantification so seductive. Numerical assessments such as indicators appeal to the desire for simple, accessible knowledge and to a basic human tendency to see the world in terms of hierarchies of reputation and status. Yet the process of translating the buzzing confusion of social life into neat categories that can be tabulated risks distorting the complexity of social phenomena. Counting things requires making them comparable, which means that they are inevitably stripped of their context, history, and meaning. Numerical knowledge is essential, yet if it is not closely connected to more qualitative forms of knowledge, it leads to oversimplification, homogenization, and the neglect of the surrounding social structure. Grounding quantitative

knowledge in a qualitative analysis of categories, meanings, and practices produces better indicators. The current rush to quantification risks sacrificing the insights of rich, ethnographic accounts.

A comparison of the information produced by quantitative and qualitative methods of studying battered women's treatment by the courts illustrates these differences. In 2006, I studied a nongovernmental organization (NGO) that did advocacy work with largely working-class and poor battered women, including African American, Caribbean, Latina, white, Asian, immigrant, lesbian, disabled, and formerly incarcerated women in New York City. These women became members of the organization. They carried out a human rights documentation project on the adequacy of New York City's family courts for battered women. Fourteen of the members, all domestic violence survivors, interviewed seventy-five other domestic violence survivors about their experiences in the courts and produced a report that outlined a series of abuses. The women who were interviewed talked about losing custody of their children to their batterers despite being the primary caretakers, about inadequate measures for safety in the court buildings, and about the unprofessional conduct of judges and lawyers that they experienced when they raised claims of domestic violence (Voices of Women Organizing Project 2008). The final report compared these problems to the standards articulated in human rights conventions.

In contrast, at about the same time, the UN Office of the High Commissioner for Human Rights (OHCHR) developed a set of indicators for measuring violence against women that were to be used by any country around the world (discussed in chapter 7). Some indicators assessed the adequacy of law enforcement in dealing with domestic violence, the same problem the domestic violence survivors in New York City were examining. The indicators measured the "proportion of formal investigations of law enforcement officials for cases of violence against women resulting in disciplinary action or prosecution" and the "proportion of new recruits to police, social work, psychology, health (doctors, nurses, and others), education (teachers) completing a core curriculum on all forms of violence against women" (UN Office of the High Commissioner for Human Rights 2012a: 99). Thus these indicators measured some dimensions of the legal treatment of domestic violence but not the problems raised by the formerly battered women in the New York City study.

Clearly, these two efforts to document battered women's experiences with legal institutions differ in the kinds of information they produced. While the first project was based on a particular local situation and generated its categories and questions from the experiences of those who went through it, the second did not address women's experiences at all. The first one used local

knowledge to decide what to count and measure, while the second relied on global measures that had already been developed and used in many different countries. The first effort took into account the ethnicity and social class of the people interviewed as well as the history of the New York City court system, while the second did not. On the other hand, its indicators allowed comparison across cultural contexts and countries in a way that the first approach did not, and it was better able to show the global size and scope of the issue.

This book focuses on the disparity between such qualitative, locally informed systems of knowledge production and more quantified systems with global reach. It argues that despite the value of numbers for exposing problems and tracking their distribution, they provide knowledge that is decontextualized, homogenized, and remote from local systems of meaning. Indicators risk producing knowledge that is partial, distorted, and misleading. Since indicators are often used for policy formation and governance, it is important to examine how they produce knowledge.

Interest in global indicators is now booming. Efforts to measure a wide variety of social phenomena took off in the mid-1990s as scholars and organizations developed indicators for such diverse issues as failed states, transparency, poverty levels, the rule of law, good governance, and the human right to health. Although indicators were developed in the mid-twentieth century to describe economic phenomena such as gross domestic product (GDP), by the end of the century, this technology was being applied to a range of social phenomena. The use of quantitative measures by national and international governments and organizations, as well as by academics and NGOs, has continued to grow in response to the demands of policy makers and the public for information about the world and as an aid to governance.

The contemporary proliferation of indicators used as a mode of governance springs, in large part, from the desire for accountability. How can states or civil society hold governments, corporations, and individuals responsible for their actions? How can donors be sure the organizations they fund accomplish what they have promised? Accountability requires information. Quantitative data, folded into simple and accessible indicators, seem ideal. Indicators of freedom, human rights compliance, trafficking in persons, and economic development are all efforts to measure country performance against global standards and to hold states accountable for their actions. Such quantitative measures promise to provide accurate information that allows policy makers, investors, government officials, and the general public to make informed decisions. The information appears to be objective, scientific, and transparent. Indicators are appealing because they claim to stand above politics, offering rational, tech-

nical knowledge that is disinterested and the product of expertise. Once indicators are established and settled, they are typically portrayed in the media as accurate descriptions of the world. They offer forms of information that satisfy the unease and anxiety of living in a complex and ultimately unknowable world. They address a desire for unambiguous knowledge, free of political bias. Statistical information can be used to legitimate political decisions as being scientific and evidence-based in a time when politics is questioned. They are buoyed up by the rise in bureaucracy and faith in solutions to problems that rely on statistical expertise. Such technocratic knowledge seems more reliable than political perspectives in generating solutions to problems, since it appears pragmatic and instrumental rather than ideological. These are the seductions of quantification.

Knowledge Effects and Governance Effects

Numbers packaged into concepts that describe social life are now central to how many people understand the world they live in. They are also central to governance. There is currently a surge of interest in systems of performance monitoring and evaluation, for example. Holding states or corporations accountable requires information on their violations. Evidence-based decision making, experimentalism, audit mechanisms, results-based management, and new public management are emerging forms of governance that rely on measurement and counting. All these forms of governance require knowledge that is classified, categorized, and arranged into hierarchies. In other words, indicators have both a knowledge effect and a governance effect.

Despite the contemporary prominence of quantified knowledge, there has been relatively little analysis of its effects on knowledge and governance. Much of the scholarship on indicators focuses on how to develop an effective, reliable, and valid measure: how to conceptualize what is to be measured, how to operationalize broad and vague concepts, what data sets are available that can be used, how to label indicators so that they will be easy to understand and use, and how to generate buy-in from governments, donors, and other potential users of the indicator. The challenges of measurement, comparability, weighting of factors, and gathering reliable data in very different historical and cultural contexts are well known and widely discussed.

My focus, however, is not on the accuracy of indicators but on the social and political processes of indicator production and their effects on regulation and governance. My ethnographic examination of the way indicators are constructed and used shows that they reflect the social and cultural worlds of the

actors and organizations that create them and the regimes of power within which they are formed. This social aspect of indicators is typically ignored in the face of trust in numbers, cultural assumptions about the objectivity of numbers, and the value of technical rationality.

Statistical knowledge is often viewed as nonpolitical by its creators and users. It flies under the radar of social and political analysis as a form of power. Yet how such numerical assessments are created, produced, cast into the world, and used has significant implications for the way the world is understood and governed. Quantitative information influences aid to developing countries, investment decisions, choices of tourist destinations, and many other decisions. A country with poor indicators for the rule of law, human rights compliance, and trafficking invites international intervention and management. Rather than objective representations of the world, such quantifications are social constructs formed through protracted social processes of consensus building and contestation. Once established and recognized, they often circulate beyond the sphere envisioned by their original creators and lose their moorings in specific methodological choices and compromises.

Beneath the “truth” of quantified knowledge, indicators are part of a regime of power based on the collection and analysis of data and their representation. It is important to see who is creating the indicators, where these people come from, and what forms of expertise they have. Rather than revealing truth, indicators create it. However, the result is not simply a fiction but a particular way of dividing up and making known one reality among many possibilities. As indicators cross the gap from social science knowledge to that used by policy makers and the public, the drawbacks and complexities recognized by their creators, such as limited data, the use of proxies, and the uncertainty of flawed or missing data, are typically stripped away. The indicators are presented as unambiguous and objective, grounded in the certainty of numbers. In this form, they act to produce a truth about the world despite the pragmatic compromises that inevitably arise in their creation. Data are never complete and may not measure exactly what the author of the indicator seeks to assess. Thus the truth of indicators can be quite misleading. For example, Morten Jerven illustrates this problem in his analysis of the flaws in information available on African economies and the impact they have on development planning (2013).

The core question of this book is how the production and use of global indicators are shaped by inequalities in power and expertise. It examines the power dimensions of indicators through an ethnographic analysis of the actors and institutions of the human rights movement engaged in the creation and use of three global indicators: indicators focused on violence against women,

indicators on trafficking in persons, and indicators of human rights violations. Through a genealogical analysis of these three global indicators, I trace the gradual process of constructing indicators from the fragments of earlier ones and the cultural assumptions and theories of social change embedded in them.

The Genealogical Method

The genealogical method asks how an indicator develops, which actors and institutions promote and finance it, and how and when its features become settled (see Halliday and Shaffer 2015). It considers how the creators grapple with converting the broad terms of a standard into a series of measurable and named phenomena. Measurement generally builds on previous models and approaches, refining or expanding them or correcting their recognized problems. Adapting existing templates and forms of data analysis and presentation requires expert knowledge, producing what I call “expertise inertia.” Expertise inertia means that insiders with skills and experience have a greater say in developing measurement systems than those without — a pattern that excludes the inexperienced and powerless. At the global level, experts are usually cosmopolitan elites with advanced education or people who have had previous experience in developing indicators of the same kind. They are often from the global North and trained in political science, economics, or statistics. Some are social scientists who research social phenomena such as political terror or violence against women.

Countries that have carried out relevant surveys create the models for the next set of surveys. The statisticians from these countries become global experts. In the context of global governance, this means that when experts gather to develop indicators and plan data collection, those from countries that have already tried such data gathering and analysis projects claim special knowledge and authority. For example, in an expert group meeting that I attended in Geneva in 2009, about twenty participants worked on developing measurements of violence against women. Representatives from Italy, Canada, and the United States talked about how such surveys had worked in their countries. People from poorer countries that had not yet carried out surveys of violence against women could not offer such authoritative expert knowledge. To understand how indicators are formed and developed, it is necessary to attend to the microprocesses through which surveys are created, categories defined, phenomena named, translations enacted. The microprocesses are, in turn, shaped by the actors, institutions, funding, and forms of expertise at play. This means that categories and models based on local knowledge are difficult to incorporate.

Those who create indicators grapple with the problem of finding or collecting data relevant to what they want to measure. Gathering data is expensive. Unless the sponsoring organization has funds to collect new data, it must locate existing data that can serve as proxies for the qualities being measured. This includes administrative data, regularly collected by governments (such as census data) or private organizations (such as electricity consumption), and social science data developed for research. Indicator creators with the resources to collect their own data may use population surveys targeted to the particular question they are interested in, but these are expensive. A cheaper alternative is the expert opinion survey. For example, instead of surveying those in the general population about their experiences of corruption, the organization can send questionnaires to local experts about the prevalence of corruption in their country. This is clearly less expensive, but also less comprehensive and accurate. Those without resources have to search out existing databases, which may not actually measure what the indicator seeks to count. The fact that existing data determine what an indicator can measure is what I call “data inertia.” It is relatively hard to address new problems without new data collection, so the way categories are created and measured often depends on what data are available.

Both of these forms of inertia inhibit new approaches to measurement and tend to exclude inexperienced and resource-poor actors from having much influence on what is measured. They relegate those with local knowledge to the sidelines. Since those who choose the template and the modes of data collection are typically powerful individuals with experience and connections to statistically advanced countries, this means that powerful and wealthy countries are likely to set the models for less powerful ones and that weaker states and nonstate actors will have difficulty influencing the shape of the indicators.

Thus it is important to track what forms of expertise are involved in creating an indicator, who pays for the experts, who funds data collection, and which organizations develop and promote the indicator. Those with experience in developing similar indicators are more often listened to and have greater influence in designing the indicator than newcomers. Local, vernacular knowledge is typically less influential than more global, technical knowledge and, based on my attendance at meetings and reading of documents, often does not enter into the discussion at all.

Temporal Dimensions of Indicator Production

The microprocesses of indicator production take place over time. Indicators and other forms of quantitative knowledge are built up through a slow, incremental process. Many are years in the making. Some of the measures

deemed most successful by the UN Statistical Commission, for example, are gross domestic product, instituted in the late 1930s, and the system of national accounts, developed first in the 1950s. Both of these measures initially required substantial theoretical work, including developing the idea that such concepts were even measurable. They also needed the creation of templates and measurement devices, mechanisms for classifying and counting, and names for the objects of measurement. They had to be presented through publicly accessible aesthetic forms and labels. Creating and maintaining indicators requires building up bodies of experts who understand them. Over time, indicators are revised as circumstances change but often remain the same in name and conception. In a few cases, indicators such as these achieve broad public acceptance. Debates continue about the details of how to measure and what to include, but the underlying concepts and measurement strategies are established.

Thus indicators gradually become more settled and less open to change. Indicator frameworks, templates, and measurements generally begin with open discussion among alternative measurement strategies and forms of data but gradually become more established and certain. This process often takes two or three decades. As the indicator crystallizes and becomes naturalized, flexible categories and proxies become fixed and unchangeable. Contestation about the indicator's underlying framework, use of data, and categories of analysis becomes more difficult over time. After a certain point, critics often succeed only in adding a variable or value. Some issues seem settled and not open to debate, while others require continuing efforts at refinement. Some of these debates concern classification and measurement, while others focus on what is to be measured and by whom. Tracing the development of indicators, their institutional basis, and the limited opportunities for their contestation and refusal reveals their quiet exercise of power.

The Ethnography of Indicators

This project is based on six years of intensive ethnographic field research that involved attending innumerable meetings and workshops, discussions with participants and others involved in global indicator projects, interviews with the major players in each of the three indicator initiatives I studied, and formal and informal meetings with activists and scholars in the United Kingdom, Europe, Australia, and India. Much of the field research depended on informal conversations with people involved in the production and use of human rights and trafficking indicators, which occurred at research meetings, academic conferences, lectures by leaders in this field in New York City and Boston, UN

events in New York City and Geneva, NGO meetings, expert group meetings, workshops for academics and practitioners, and treaty body meetings and during my travels to lecture or attend conferences. I talked to academics, statisticians, human rights activists, international lawyers, and people who work for international NGOs and the UN.

In addition to ethnographic information, this book relies on the extensive documentary record available for such activities, including both the documents of meetings I was unable to attend and the records of quantification projects. UN documentation is particularly rich, and although it does not describe the informal negotiations behind the documents, it does offer a wealth of formal information, which I have supplemented by attending meetings and conferences and talking to the principal actors in these processes. My work was supplemented by fieldwork by my graduate students, Jessica Shimmin in the United States, Vibhuti Ramachandran in India, and Summer Wood in Tanzania. This is a study not of a particular place but of a global one: it traces processes that stretch across nations and continents. This is a transnational, deterritorialized social space, rich with shared meanings, practices, and technologies. I refer more specifically to these meetings and interviews in the chapters that follow as well as to the range of documentary evidence I consulted.

Governance and Indicators

THE EMERGENCE OF “INDICATOR CULTURE”

The increasing importance of quantification in governance reflects the emergence of what might be called an “indicator culture.” It is a dimension of what has been labeled audit culture (Power 1999; Shore and Wright 2015; Strathern 2000). I use the concept of culture to refer to a set of techniques and practices applied within specific situations rather than as a description of a society. Thus it is a set of cultural practices, techniques, and assumptions about knowledge production embedded in particular institutional and bureaucratic settings. It is a culture in the sense of Shore and Wright’s discussion of audit culture: it is not a holistic set of actions and ideas that define a society but a technology that occurs in a variety of contexts (2015). It is part of the repertoire of institutional actors seeking to persuade publics and influence governance decisions. “Indicator culture,” in this sense, includes a body of technocratic expertise that places a high value on numerical data as a form of knowledge and as a basis for decision making. Its characteristics are trust in technical rationality, in the legibility of the social world through measurement and statistics,

and in the capacity of numbers to render different social worlds commensurable. This perspective includes a pragmatic acceptance of imperfect measurement and skepticism about politics. It builds on social science expertise and its claims to objectivity for credibility and legitimacy. It assumes that all things can be measured and that those measures provide an ideal guide to decision making. Adherents of this approach to governance see data as the basis for policy and audit mechanisms as essential for management. The use of indicators for governance depends on the belief that experts can generate commensurable knowledge across substantial differences in language, culture, history, and place. From this perspective, indicators enable policy makers to compare freedom in Mauritius and Mauritania, poverty in Sweden and the Sudan, and human rights compliance in Russia and Rwanda despite the vast differences between these countries.

Indicators are often employed as a technology of governance in situations where lines of authority are unclear, law is soft rather than hard, jurisdiction is ambiguous, and governance requires negotiations among sovereign nation-states. They are important in regulatory situations that are governed by guidelines rather than rules. Development economics, development aid, public health, international trade and investment, global educational systems, and human rights monitoring are only a few of the areas where quantitative information is increasingly fundamental to decision making. In all these areas, universal standards have developed to which states are held accountable so that indicators can be used to make these standards more specific.

EVIDENCE-BASED GOVERNANCE

The rise of indicator culture is connected to an emerging mode of governance that is referred to as “evidence-based governance” or simply “new governance,” a broad range of regulatory strategies that rely on empiricism, quantitative knowledge as the basis for decision making, and problem solving through benchmarking (de Burca 2010; Power 1999; Rittich 2014: 175). Key features are broadly framed goals, stakeholder participation, flexibility, reversibility, monitoring and peer review, transparency, a data-based approach, and learning-oriented and multilevel decision making (de Burca 2010: 235). It is related to “new public management,” often referred to as “results-based management,” and represents the movement of business management techniques to the public sector. It is essentially a shift from a command-and-control strategy of governance to collaborative, consensus-building discussions focused on problem solving and improvement (Simon 2004: 11–28). This model encourages learning

and innovation, constant collaborative revision, and the participation of multiple stakeholders (Simon 2004; de Burca 2010; Rose 1991).

The state is no longer the only player, since nonstate actors also participate in formulating laws and policies and in the development and enforcement of norms at the international, as well as the domestic, level (Thomas in Halley et al. 2006: 385). In this regime, the use of evidence organized by guidelines, standards, metrics, and performance evaluations is essential to decision making. Audits are used to determine the quality of data. In contrast to earlier systems, which relied on rules and punishments for violations, this mode of governance works through the collaborative production of standards and the evaluation of outcomes, including the use of self-assessment and ranking techniques.

Evidence-based governance relies extensively on “soft law” in that it shapes behavior by establishing standards and requiring individuals, groups, corporations, and even nations to report on how they have met these standards (de Burca 2010; Trubek and Trubek 2005). Instead of imposing sanctions, as “hard” law does, it seeks enact change through assessment, reporting, and ranking. For example, countries that fail to meet targets or that are ranked below others on key indicators are to be “shamed” into improving their records (see also Maurer 2005; Trubek and Trubek 2005). Evidence-based governance requires information on performance assessed with reference to standards. For example, compliance with standards for the right to health is measured by maternal mortality and life expectancy. Evidence does not lead to the imposition of sanctions, but to correction and advice about how to improve performance. This approach is now widespread in such fields as development and human rights compliance as well as in many other domains of global and local governance.

Governance by indicators can increase egalitarian decision making and accountability by opening up the basis for decisions to public scrutiny. On the other hand, it can also reinforce inequality and evoke resistance among the governed. For example, Marilyn Strathern and her colleagues criticize the university evaluation program of the British government, which has introduced indicators of faculty productivity and activity as the basis for allocating revenues to academic departments (2000; Shore and Wright 1999; 2000). This mechanism creates standards that universities and professors are responsible for satisfying. These professors — the governed — argue that this regime displays a lack of trust in the faculty and leads to alienation, exhaustion, and withdrawal.

With the turn to evidence-based governance, responsibility for decision making is shifted from individual, discretionary judgment to systems of measurement established by experts. Indicators displace the capacity for judgment from those assessing performance to the creators of indicators used to assess

performance. This shift reduces discretion and private decision making, opening governance up to greater public scrutiny. At the same time, it moves responsibility from judicial and political decision makers to the experts in quantification who develop and implement measurement systems. Ultimately, indicators place responsibility on the governed to conform to the indicators, regardless of who has created them.

Defining Indicators

What is an indicator? In practice, this is a very broad and vague term that refers to a wide range of quantitative and qualitative techniques for ordering knowledge. It grows out of the basic idea of signs or markers that have a particular meaning. For example, one could specify behaviors that *indicate* that a person is drunk. However, in the field of global governance, the term is used to refer to a variety of approaches to packaging knowledge. Davis, Kingsbury, and Merry (2012) define indicators as follows: “An indicator is a named collection of rank-ordered data that purports to represent the past or projected performance of different units. The data are generated through a process that simplifies raw data about a complex social phenomenon. The data, in this simplified and processed form, are capable of being used to compare particular units of analysis (such as countries or institutions or corporations), synchronically or over time, and to evaluate their performance by reference to one or more standards” (2012: 73–74).

Indicators refer to the systematic, comparative organization and presentation of information that allows for comparison among units or over time. Indicators create and define social phenomena by naming them and attaching them to data (see Davis, Kingsbury, and Merry 2012). When an indicator is labeled, it defines the phenomenon it is measuring. For example, it is hard to define intelligence, but the concept is often specified by what the IQ test measures. The process of measurement tends to produce the phenomenon it claims to measure. An indicator is labeled as measuring, for example, rule of law or corruption. It then specifies a series of measures that constitute this concept. It defines the concept by linking it to specific criteria and measurements.

Indicators are different from targets and goals, which specify objectives. Indicators provide information that can be used to assess compliance with targets and goals. They attach data to a standard in order to assess performance against that standard. For example, the Millennium Development Goals (MDGs) have goals and targets with indicators attached to each target. In reference to women’s rights, Goal 3 of the MDGs specifies, “Promote gender

equality and empower women.” This is followed by Target 3A, “Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015,” which is measured by Indicator 3.1, “Ratios of girls to boys in primary, secondary and tertiary education,” Indicator 3.2, “Share of women in wage employment in the non-agricultural sector,” and Indicator 3.3, “Proportion of seats held by women in national parliament” (United Nations 2008a). Finally, a “benchmark” describes a data point that constitutes a goal to be achieved.

In practice, the term “indicator” is used for many different kinds of numerical representation with no real consistency. Meanings range from simply something that indicates or points to a fact, such as that a person has been trafficked, to an elaborate combination of data merged into a single rank or score. Todd Landman defines an indicator as “a distilled measure of a concept” (2010: 137, n. 1). He points out that an indicator may not be able to represent the totality of a concept, particularly if it has multiple elements and dimensions. It is the simplification of information, the extraction and classification of some diagnostic element out of the buzzing array of particular features of the social world, that is the hallmark of indicators.

Indicators vary in the extent to which they incorporate qualitative information, local knowledge, and contextual data. As this book indicates, there are debates about the construction and use of indicators, including the extent to which they should reflect national and local knowledge. Some indicators do a better job of representing social life than others. Those that are more accurate tend to concentrate on phenomena that are readily countable, that require less interpretation, that employ qualitative research to generate the relevant categories for counting and analysis, and that are able to gather data that are appropriate to the measure rather than relying on proxies and fillers for missing data. Indicators that offer more complex and multifaceted measures that are less superficial may suffer in the competitive marketplace of indicators, where simplicity, ranking, and conformity to popular beliefs foster indicators’ acceptance.

TYPES OF INDICATORS

Indicators can be used for advocacy, monitoring, and social science scholarship. Advocates use indicators to make problems visible, social science scholars to produce scientific knowledge, and businesses and governments to monitor and control behavior. Contestation over whether statistics are to be used for advocacy, management, or scientific knowledge stretches back at least to the begin-

nings of government statistics in the nineteenth century and recurs in contemporary debates about indicator production. In a meeting I attended charged with developing measures of violence against women, for example, some of the participants were experts in research and statistics while others worked in UN offices or for NGOs. The group confronted the difficulty of deciding whether the goal of an indicator is to develop a scientific descriptive tool for theory development, a policy tool to assess government policies and NGO programs, or an advocacy tool to shame governments for poor performance. These are three quite different objectives. The first is to produce scientific knowledge, the second is to facilitate organizational management, and the third is to promote reform. Each suggests a somewhat different approach to formulating questions and selecting items.

Within the field of global governance, it is also possible to distinguish three types of indicators on the basis of their mode of quantification. These are based on Michael Power's analysis of first- and second-order measurements (1999), but in this context, it is necessary to add a third order. Much of the recent growth in indicators has occurred in the third order. Third-order indicators are constructs put together from other indicators, usually building on the knowledge created by first- and second-order measurements. The distinctions are not sharp; there are continuities among them.

Counts. Counts refer to numbers of people, things, events, or laws. Census data are a prime example, as are the results of many kinds of surveys, such as health and demographic surveys, crime victimization surveys, and opinion surveys. Surveys may involve a sample, or they may seek complete coverage, as a census does. Even simple counting raises three questions: (1) What is important to count? (2) What characteristics are diagnostic for identifying these countable things? (What, for example, should constitute violence against women, or how should caste identity be defined?) (3) What are the appropriate criteria for aggregation and disaggregation? Counts require cultural work: they depend on constructing categories such as gender, ethnicity, income, and employment status. Creating categories implies deciding on where to lump and split, what to include and what to leave out, how many categories to use, and what the criteria for these categories should be. The process is a deeply interpretive one, reflecting the major preoccupations of a society (Bowker and Star 1999). Countries count what they care about. For example, the US Census counts race but not religion, while Denmark does not count race; these differences reflect national histories.

There is interpretive work in determining how categories are constructed

and where to disaggregate. Is average income data adequate, or is it necessary to disaggregate by race, by class, by urban versus rural residence? How are these categories to be defined across countries? Is gender a binary or a continuum? How should race be defined? Many categories must be constructed, such as those of race, unemployment, poverty, and even age. Surveys and censuses often count households, but what constitutes a household? Once these categories are created, there is a tendency for them to remain stable — the product of the pragmatics of data collection and “the inescapable inertia of categories already in use” (Bowker and Star 1999: 117). Changing categories requires a new regime of data collection and undermines comparison over time.

Ratios. Ratios, which compare two numbers, facilitate comparison among countries or organizations. Indicators of this kind evaluate one number against another. Instead of providing the number of people who are unemployed, for example, a ratio indicator describes the rate of unemployment per capita in a population. However, ratios also require the creation of categories and raise similar questions about lumping and splitting. It is also necessary to determine the baseline of the comparison. Should the rate of unemployment be calculated against employed persons? Or the total population? Or the population of working-age persons? Should they be calculated against only men or both men and women? Similarly, should the rate of domestic violence be counted against all women? Or all married women? Or all women who were ever in a partnered relationship? There are clearly many critical decisions that affect the final number. Ratios are widely used, including those of life expectancy, per-capita income, maternal mortality, and poverty. They are much more readily comparable across countries and over time than counts. Many of these ratio indicators are well established and accepted.

Composites. Composite indicators are the most widely known, persuasive, and referenced kind of indicator. Composite indicators are made up of separate systems of counts and ratios, often merged together. They combine multiple sources of data, even multiple kinds of data, converted into a single score or rank. They also merge different attributes into a single measure. They are, in effect, a basket of counts and ratios combined to define a single concept. They usually combine several measurement systems and weight the constituent parts to construct a measure of a complicated idea such as rule of law or corruption. Composites require significant interpretive work in naming, weighting, and combining elements. They may have multiple dimensions. For instance, the Human Development Index has three dimensions, the World Bank Governance Indicators have six, and the MDGs have forty-eight. They typically

rely on data from other indicators. Naming a composite indicator is important, since the name defines the concept. In effect, composite indicators construct a concept by naming and measuring it.

All indicators are framed by implicit theories about what is important to count and what is not, as well how social change happens, but theoretical elaboration is greater with composite indicators than with counts or ratios. Composites crystallize complex theories such as how the rule of law works and what its constituent elements are, what constitutes a failed state, and what laws and policies promote human rights. Interpretation of composites is more extensive than that of counts or ratios because composites require merging and weighting. Composite indicators are farther from the underlying data than either counts or ratios, and it is harder to trace back the process through which behavior is converted into data and then into the indicator. They often use pre-existing data sets, which serve as proxies for the variables, producing long interpretive chains. Some are presented as ranks, some as tiers, and some as scales. In a ranking system, each entity is compared to all the others in a hierarchy so that the improvement of one entity inevitably means the decline of another. Scale and tier systems lack this zero-sum quality. For example, Freedom House rates every country on a scale of “free,” “partially free,” or “not free.” A tier system divides the population into layers or tiers based on some measurable criterion. Ranking systems tend to be the most influential indicators but also the ones that evoke the most resistance and complaint, particularly from those who are ranked poorly (see Cooley and Snyder 2015).

SUCCESSFUL INDICATORS

International organizations, governments, NGOs, academics, and UN agencies continually generate new indicators (see Davis, Fisher, Kingsbury, and Merry 2012; Merry, Davis, and Kingsbury 2015). The ecology of indicators is dense, with multiple competing measurement systems for issues such as failed states or the rule of law. A few of these become widely accepted and used, while the vast majority of indicators search in vain for global interest and influence. Successful ones are routinely cited in the media, disseminated to a wider public and gradually accepted as more or less accurate descriptions of the world. Successful indicators tend to be composites that are relatively simple in conceptualization, that are developed and promoted by powerful actors and organizations along with networks of supporters in the United Nations and governments, and that rely on academic expertise from prominent universities. Composite indicators catch the attention of the media and the public, since

they summarize a large amount of information, allow for comparison and ranking of units such as colleges and countries, and offer shorthand knowledge of complex situations. Indicators have greater credibility if their rankings conform to widely accepted views of good and bad performance among countries. They usually have a long trajectory of development, often stretching over twenty or thirty years.

When an indicator is successful, the indicator and the theory embedded in it enhance each other's popularity. An index can promote a well-established idea, as in the case of Freedom House, which measures "freedom in the world" using the concept of Western liberalism (Bradley 2015), or it can introduce a new concept, as in the case of the Human Development Index (HDI), which replaced per-capita income with a broader measure of human well-being. Converting a theoretical idea, such as the nature of the rule of law or the prevalence of modern-day slavery, into an index that ranks countries aids its dissemination.

Some of the most successful versions of composite indicators seek to measure features of governance such as corruption, freedom, and the rule of law. One example is the Corruption Perceptions Index developed by a Berlin-based NGO, Transparency International. Originally designed to bring greater focus to the issue of corruption for international development, Transparency International now works through a variety of national offices to produce comparative measures of the level of perceived corruption in almost all countries of the world. One of the oldest indicators is Freedom House, a US-based NGO that began in the 1970s and now publishes an annual report, *Freedom in the World*, assessing the degree of freedom in most countries of the world (Bradley 2015). The Global Reporting Initiative, based in the Netherlands, is a multi-stakeholder NGO that works with businesses and civil society organizations to develop indicators for corporate social responsibility (Sarfaty 2015). A relative newcomer to the global measurement project, the privately funded World Justice Project (WJP), based in Washington, DC, publishes a Rule of Law Index assessing the extent to which countries adhere to the WJP's principles of the rule of law.

Other prominent global indicators include the Doing Business Index of the International Finance Corporation of the World Bank, the World Bank-supported Worldwide Governance Indicators, the World Bank's Country Performance Institutional Assessment (CPIA), and the United States' Millennium Challenge Corporation indicator, used to determine which countries are eligible to receive certain US aid funds (see further Merry, Davis, and Kingsbury 2015). An influential composite indicator of human trafficking is the US State Department's *Trafficking in Persons (TIP) Report*, the subject of chapters 5 and 6.

In the field of economic development, the Millennium Development Goals (MDGs)—a set of targets and goals established by the UN, based on the Millennium Declaration of 2000—have been extremely influential. Created in the early 2000s with the goal being realized by 2015, they are now being revised for a new, post-2015 version.

Another widely accepted global indicator of economic development is the HDI. The HDI is popular because it is simple, straightforward, easy to understand, and promoted by a powerful international development organization, the UN Development Programme (UNDP). It articulates a new theory of development through its index. Developed in 1990 to replace the use of only gross national product (GNP) per capita as the measure of development, the HDI expresses the theory that social and economic development are inextricably related and need to be considered together. Instead of focusing on economic growth by itself to measure development, this indicator combines economic and social factors in what is called a “capabilities approach” that emphasizes ends, like a decent standard of living, over means, like income per capita. Following Amartya Sen’s capabilities approach, it measures access to health, education, and goods that give individuals the capacity to achieve their desired state of being (Stanton 2007: 3; Sen 1999; Sen 2005). This approach constituted a new understanding of development itself. As a recent study observes, “In 1990, the United Nations Development Program (UNDP) transformed the landscape of development theory, measurement, and policy with the publication of its first annual *Human Development Report (HDR)* and the introduction of the Human Development Index” (Stanton 2007: 3).

The history of the HDI reveals several key features of a successful indicator. First, it is the product of a long period of research, analysis, and experiment. Second, it is promoted by a leading, powerful institution and formulated by development economists and international policymakers located within prominent academic and policy centers. Third, it expresses, but does not test, a theoretical position. Indicators typically embody, but do not explicitly articulate, a theory of social change. Fourth, it is recognized by its creators as a very simplified representation of a far more complex body of data but is promoted for policy makers who want a convenient and quick summary. Indicators are quite distinct from the underlying statistical data that constitute them since they are single numbers or ranks designed for ease of comprehension and use as well as accuracy. Fifth, it is politically acceptable. Although there has been considerable debate and controversy over the HDI, it has become established while other indicators developed by the UNDP, such as a political freedom index, have not.

Composite indicators that rank countries tend to be particularly influential. They often present their rankings through color-coded maps, typically coloring top countries green and bottom ones red. The HDI, the Transparency International Corruption Perceptions Index, the World Bank Institute Global Governance Indicators, the US State Department *Trafficking in Persons Reports*, and the Freedom House *Freedom in the World* report all use some form of comparison and ranking. Most present their findings using color-coded maps.

A successful indicator is built up over time and gradually acquires credibility and the appearance of objectivity and truth. To achieve public credibility, it needs strong institutional support and an appealing underlying theory. Indicators that reinforce existing ideas about good and bad countries according to the relevant criteria fare better in the competition for attention and influence. If it is obvious that data collection is thin or inaccurate, it will undermine the credibility of an indicator, even though the empirical basis for the indicator is usually opaque or presented in a sketchy way.

However, even successful indicators face political challenges. Those that are the most influential, that use radical forms of simplification, and that permit scoring and ranking of countries tend to be the ones countries resist most vehemently. For example, countries resist being ranked according to their compliance with human rights standards. Many countries resent the unilateral ranking system of the US State Department's annual *Trafficking in Persons Reports* (see Gallagher and Chuang 2012). Indicator systems have to work around sovereignty concerns. Some of the most successful indicators are embroiled in battles over their data and methodologies from time to time despite widespread agreement about the general structure of the indicator. For example, although it is a prominent, widely used measure, there have been ongoing controversies over the HDI (Ward 2004: 200–203). An intense debate about the latest HDI focusing on its procedures and the data it used roiled the UN Statistical Commission in 2011. The debate also showed how important the HDI is to many countries around the world (Merry 2014).

THE MYTH OF OBJECTIVITY

Indicators promise to provide objective knowledge but sometimes fail in at least two ways. The first is creating false specificity: they appear more accurate and precise than they are. Exact rankings of countries, precise numbers of trafficked victims, and percentages of women who have experienced violence presume that these are countable phenomena that can be compared and added together. The ambiguity of the categories, errors in counting, missing data, and

lack of commensurability disappear in the final presentation of the indicator to the public. In order for an indicator to succeed in policy and public domains, it must present information in a simple and unambiguous way without a great array of qualification and methodological discussion. Indicators endeavor to persuade within a rich ecology of competing indicators, in which the simplest and most coherent often prevail.

The second possible reason indicators fail to produce objective knowledge is that they camouflage the political considerations that shape the collection and presentation of data. By “political,” I do not mean the kinds of pressures political leaders exert on official statistical bureaus to produce the data they wish. My analysis of the politics of indicator production refers to the ways in which indicators are subtly and even unconsciously shaped by the assumptions, motivations, and concerns of those who carry them out. Since indicators are produced by individuals, networks, and institutions with their own interests and agendas, the producers’ perspectives shape the outcome. Indicators produced by advocacy organizations are more explicit in their agendas, such as showing that there is a large population of trafficking victims, but indicators that claim to produce unbiased data also reflect particular interests and perspectives. They are shaped by the disciplinary and institutional site of their creation and by the resources available to collect relevant data.

Since states and private actors often rely on indicators to make policy decisions and promote state accountability, it is urgent to examine how they are developed and how they work. By recognizing the politics of producing quantitative data, it is possible to see how particular choices about how to categorize and count shape the knowledge that is produced as well as what is missed, ignored, and not counted. For example, the US State Department’s *Trafficking in Persons Reports* expose deficiencies in state efforts to prosecute traffickers, but they do not consider the failure of states to tackle rural poverty or oppressive marriage practices, even though both of these also fuel trafficking.

Indeed, despite claims to objectivity and transparency, indicators are built on a string of interpretive decisions. Although they rely on quantitative information that appears unambiguous because it is numerical, interpretations creep into the final product at each step along the way. The choice of measurement approaches, the construction of categories, the selection of data sources, the use of proxies to measure a concept when specific data are unavailable, and the label used for the phenomenon that is being measured are all matters of choice and interpretation. They define what the concept is, how it is understood, and what things can be counted to measure it.

However, this is not a simple story of a hegemonic global technology

imposed on passive and helpless local communities. Any global concentration of knowledge depends on practices of counting and measuring within countries and communities around the world. International statistics are often patched together from national data sets along with those from NGO, county, city, or regional data collection systems. They rarely fit together easily but must be massaged and made to fit through statistical techniques. Classification systems often grow out of local systems of knowledge that become globalized systems into which the local systems of other countries must be squeezed. To stitch together local systems of classification, it is necessary to find ways to make different things commensurable (Bowker and Star 1999). For example, in order to measure violence against women, throwing acid in the face of one's wife in Bangladesh must be equilibrated to shooting a domestic partner in the United States. This intellectual, interpretive work is shaped by the politics of expertise and participation that determine how quantitative knowledge is developed and by whom.

In the end, those who create indicators aspire to measure the world but, in practice, create the world they are measuring. In other words, indicators do not stand outside regimes of power and governance but exist within them, both in their creation and in their ongoing functioning. They are a blending of science and politics, of technical expertise and political influence. The two work hand in hand, sometimes in overlapping or competitive ways, with considerable slippage between them. The technical is always political because there is always interpretation and judgment in systems of classification, in the choice of things to measure, in the weighting of constitutive elements, and in decisions about which denominator to use for a ratio. The political hides behind the technical. Technical knowledge may be used to avoid political discussion, to cover up or legitimate political decisions, or to displace responsibility for decisions. Indeed, technical experts typically conceive of their work as, ideally, outside the domain of the political.

Clearly, this analysis of quantification grapples with a central epistemological problem in social science: the relationship between quantitative and qualitative methodologies in producing knowledge. As generations of scholars have argued, each provides insights, but not alone. There are inevitable trade-offs. Qualitative knowledge such as a detailed ethnography of a village by itself fails to examine to what extent this village represents any larger society. Quantitative knowledge alone inevitably selects a few features for comparison, ignoring their specific histories, interconnections, and locatedness. It lacks a holistic perspective and has difficulty embedding the analysis in a social context. Both methods of research taken separately contain hazards.

Narratives taken out of context can be misleading, just as numbers without context are. As Haltom and McCann point out, moralistic narratives that simplify tort cases present distorted pictures of the way law works and may lead to policy changes. For example, the account of a woman burned by McDonald's hot coffee in 1992 who sued the company and received a large damage award was widely circulated in the media, contributing to a myth of litigiousness and to changes in the way litigants, judges, lawyers, and the general public thought about torts and their willingness to complain about violations. Yet the media focused on the large award and dropped many of the important features of her case, such as the extent and severity of her burns and the existence of a long string of similar complaints in the past that had forced the company to acknowledge the problem. Instead, media stories focused on the litigiousness of the plaintiff (Haltom and McCann 2004: 191–226). These stories contributed to the tort reform campaign by claiming that the case was outrageous and unjustified. Haltom and McCann argue that the use of such truncated narratives by entertainment-focused media creates a public common sense about lawsuits that inhibits deeper discussions about public policy, which is ultimately undemocratic (Haltom and McCann 2004: 24). Both stripped-down narratives and stripped-down numbers can provide the basis for popular moralistic accounts that conceal and distort the dynamics of power and obstruct public debate.

Overview of the Book

This book provides a genealogy of three indicators in the context of a wider set of indicator practices and their historical origins. Each indicator has a different institutional sponsor, resource base, and form of international collaboration. They raise different issues about knowledge production and governance, translation and commensuration, and the challenges of presenting knowledge through numbers. Yet in all three cases, numbers are used to strengthen narratives and to persuade audiences of the validity of their underlying theoretical argument. In all three, the process of development is slow and builds on expertise, past experience in measurement, and social science quantification techniques. And all three resist efforts to challenge or change them.

Chapter 2 presents the theoretical framework for the project, using the power/knowledge framework and science and technology studies to examine the political dimensions of the production of quantitative knowledge. It also provides a historical context for the development of statistics, and their use in both national and colonial governance, and traces the formation and devel-

opment of an international statistical body, the UN Statistical Commission. Chapter 3 describes a project to develop global indicators for violence against women by the UN Statistical Commission. A major challenge of the project was developing categories for analysis that rendered the diversity of the phenomenon commensurable. A global survey of violence against women requires shared understandings, comparable categories of measurement, and some consensus on what violence against women means. My ethnography of the process revealed several parallel initiatives to measure violence against women with different theoretical frameworks and measurement strategies. Chapter 4 discusses four of them: the gender equality approach, the human rights approach, the criminal justice approach, and the national statistical approach. Tracing the effort to produce global data on violence against women shows tensions between feminists and statisticians and global, rather than local, definitions of the problem. It also reveals the critical role played by social science expertise and institutional support.

Chapter 5 examines the US State Department's *Trafficking in Persons Reports*, published annually since 2001. They assess countries' performances in combating trafficking in persons based on a US State Department survey and assessment of antitrafficking efforts. The result is a system of ranking, carried out since 2001, that uses tiers rather than numerical scores. It ranks countries according to their compliance with a set of standards for combating trafficking, developed by the United States. It is authorized by the US Congress as a way to diminish trafficking from source countries. This indicator system is a unilateral exercise by one country that evaluates performance according to its own standards. Its intellectual groundings are largely in the field of criminal justice and prosecution, and the framework parallels an earlier ranking of countries in terms of their efforts to control narcotics. The measurement system is framed by US foreign policy, promoted by the US secretary of state, and based on data collected by US embassies around the world. Chapter 6 compares the quantitative approach to understanding and governing trafficking with the one provided by ethnographic studies.

Chapter 7 analyzes a system for measuring human rights compliance developed over seven years by the UN OHCHR. This indicator emerged out of a thirty-year effort to develop human rights indicators for social and economic rights, such as the right to health and the right to food, and to convert the broad legal obligations of human rights conventions into more specific commitments for states. The OHCHR's project generated indicators for twelve core human rights, such as the right to liberty and the right to health, and two cross-cutting rights: nondiscrimination and violence against women. These indicators are

designed to support human rights monitoring by treaty bodies, the committees that oversee compliance with human rights treaties. While each human right is measured by a variety of indicators, the project does not convert these indicators into a single measure nor does it engage in ranking. This may explain its relatively slow adoption. The chapter shows how indicators serve to translate knowledge from one domain, law, to another, development economics. However, this transformation undermines the legal dimensions of human rights in order to make its ideas more accessible to wider policy and development audiences.

All three of these indicator projects confront several dilemmas. One is the need to create measures that can be compared with other countries yet are tailored to the conditions of a particular country. This is the basic dilemma of creating commensurability while maintaining flexibility. It is necessary to develop fixed categories across organizations and over time in order to compare and evaluate change, but these categories need to be defined differently to account for variations in practice, context, and history. A second dilemma is acquiring relevant data. Either data must be gathered in the categories specified by the indicator, which is a very costly process, or indicators must use existing data as proxies for what they attempt to measure. The use of proxies is problematic if they do not cover the issues the indicator seeks to measure. For example, if a study wants to assess the relative burden of water gathering on men and women, it will not be easy to use data from a household survey that simply asks if the household has access to water.

A third dilemma is promoting acceptance of indicators by policy makers and publics. Indicators that conform to existing conceptions of the world or established theories are more likely to be accepted than those that promote a new theory or provide a different way of ordering the world. Thus indicators that present new ideas may lose out in the marketplace of indicators. A fourth dilemma is the need to simplify information in order to enhance general acceptance despite the desirability of creating more complex and disaggregated categories of data and an analysis that offers a more accurate picture of social phenomena. A fifth dilemma is the tension between technical expertise and policy concerns about what, where, and how social phenomena can be measured. This dilemma pits statisticians against politicians, with politicians asking statisticians to measure phenomena that are vague, illegal, or politically motivated and statisticians insisting on their autonomy and professional judgments about what can be measured accurately.

The genealogy of these three indicators shows how the participants, organi-

zations, and communities of expertise involved in their production and dissemination manage these dilemmas. This includes looking at the histories of each indicator and the templates and categories that they adopt. There are moments of political contestation and public debate, as well as private discussions and expert meetings, in the development of each one. Like scientific knowledge, some indicators gradually acquire certainty over time and win the support of an expanding network of experts, while others lack this support and disappear. At first, measurement systems are open and experimental, but over time, a more settled knowledge emerges about how to measure things. While tweaks to frameworks may be accepted, major shifts in frameworks and measurements are often resisted.

Conclusions

Indicators emerge through social processes shaped by power relations, expertise, and techniques of measurement. The statisticians and experts who create indicators confront challenges of missing data and unmeasurable phenomena, but through pragmatic compromises, they manage to produce quantitative knowledge that shapes public attitudes and policy decisions and responds to the human desire to know the complicated and often unknowable world. Given the increasing use of indicators, it is important to interrogate these forms of knowledge and their limitations as well as to keep open channels of contestation and resistance to their hegemony. Those who are measured typically lack a voice in the construction of the categories and measurements. Moreover, subordinated groups have difficulty resisting or changing indicators. They may be able to tweak the measurement system but rarely have the opportunity to fundamentally restructure it. Since indicators typically develop over a period of time, they become progressively harder to change. If an indicator survives competition with other indicators and is widely accepted, it comes to provide a kind of unassailable truth.

Yet indicators are not all the same. Some provide a more accurate and complex understanding of social phenomena than others. As the case studies in this book indicate, some are more superficial and simplified than others, some rely more on qualitative data in constituting their categories and analyses, and some are more attentive to local and regional conditions than others. Some are locally generated to bring attention to a problem in terms of numbers as well as stories, while others are globally produced to bring recalcitrant nations to heel. Counts and ratios stay closer to the underlying thing they are measuring, while

composites have longer interpretive chains between the counting and the final presentation of information. They are more accessible and more problematic, easier to use and less transparent.

Given the power of quantitative knowledge, it is important to increase indicator literacy both to understand the strengths and the limitations of quantitative knowledge and to compare and assess different indicators. The goal of the book is to develop a more skeptical view about indicators and to provide criteria for assessing their relative merits. Like other forms of quantitative and qualitative knowledge, indicators emerge from a regime of power relations and interpretive work. Some are more transparent than others. Some rely more extensively on qualitative research to develop categories of analysis than others. There is no doubt that such forms of knowledge provide a more reliable basis for decision making than ignorance or prejudice, but it is important to balance numerical knowledge with the qualitative knowledge provided by ethnography, human rights documentation, and qualitative interview research.

By opening up the social processes by which they are formed and the underlying theoretical and political interests of those who develop them, this book seeks to make indicators' particular ideological and structural biases more visible. The case studies in the following chapters examine the complexities of this process of indicator production, the uneasy compromises that designers are forced to make, and the variety of indicators that they produce. Understanding what numbers do and do not say and the politics underlying their creation challenges the seduction of quantification: the idea that numerical data offer a particularly reliable form of truth.

Indicators as a Technology of Knowledge

Indicators are a technology of knowledge creation, one that depends on processes of translation and commensuration. Creating indicators requires translating social life into commensurable categories so that different events become instances of the same thing. This requires deciding how to convert the wide array of practices, structures, and political and economic systems that make up social life into commensurable categories for enumeration. As Espeland and her coauthors show, commensuration requires substantial reframing and cultural work (Espeland and Sauder 2007; Espeland and Stevens 1998; see also Comaroff and Comaroff 2006). Making things commensurable depends on identifying a core principle that they all share and that renders them various instances of the same thing. It means specifying points of similarity and ignoring other features, unbundling the whole entity under consideration into discrete, countable parts. Categories must be distinct enough that cases can relatively easily be assigned to one or another. The process of categorization inevitably constricts the way that social action is understood. For example, counts of domestic violence cannot include the kinship networks, gender norms, attitudes toward violence, or history of the relationship of a particular person. Yet it is these factors that determine the way a person experiences domestic violence. The process of translation homogenizes populations, actions, and practices and strips them of context. Sometimes this means defining an act in a way that diverges radically from the way it is experienced. For example, in order to count victims of sex trafficking, it is necessary to merge a wide variety of paths through which women become involved in sex work with