

Measuring the Impact of Human Rights: Conceptual and Methodological Debates

Christopher J. Fariss
University of Michigan
cjfariss@umich.edu

Geoff Dancy
Tulane University
gdancy@tulane.edu

Fifty years ago, the world had very few human rights laws, very little information on human rights violations, and very few indicators of those violations. Today, the situation could not be more different. Though law did not directly cause the production of indicators of human rights and law, it has indirectly influenced how observers and academics use information to generate, develop, and change those indicators. Two perspectives have developed to explain these trends. The factualist approach measures whatever information is available, however imperfectly, and assumes that the resulting indicators are valid representations of the theoretical concepts of interest. The constructivist approach reminds us that these processes are not independent, and that a science of law and human rights is fallible. Though the conclusions from these perspectives diverge radically, they agree on a central notion: that international human rights law has contributed very little to social progress. We disagree and offer an alternative, constitutive approach that both accepts the critique of indicators and offers a way forward that encourages scholars to treat measurement itself as an object of theorizing and inquiry. Moreover, the approach provides scholars with guidance on how to link theoretical insights about the relationship between law, human rights, and information production with new theoretically informed measurement models that are designed to accommodate these ideas by making use of the flawed and incomplete pieces of information that are currently available. The empirical pattern that emerges from new studies that embrace this approach suggests that not only is there a relationship between law and data collection, but also, once this relationship is taken into account, that law has promoted improvements in the human condition.

Forthcoming in *Annual Review of Law and Social Science* 13:TBD (November 2017)

I. Introduction

Binding human rights laws and human rights indicators emerged simultaneously over the last fifty years. As of December 16, 1966, three total multilateral human rights treaties were opened for adoption, and no comprehensive effort existed to measure rights-based practices. Observers struggled to count and catalogue atrocities in European dictatorships, war-torn Vietnam, and apartheid South Africa (Benenson 1961, Norden 1966, The Anti-Apartheid Movement 1966). By 2016, nine multilateral treaties, nine optional protocols, three regional bodies of law, and 99 additional international instruments protect over 60 human rights.¹ These laws are now accompanied by an ever-increasing number of indicators. To date, roughly 400 quantitative and qualitative measures of human rights, democracy, and rule of law that have been coded by projects making use of standardized human rights documents, journalistic accounts, the historical record, and most recently thousands of collaborating experts.² Observing these historical developments, one may justifiably assume that human rights law caused indicators to come into being. After all, law demands the standardized collection of hard evidence, and at a deeper level, both law-making and statistics owe their existence to the enduring human desire for rationally ordered society (Dezalay & Garth 2002, Scott 1998).

However, human rights law and human rights indicators were neither directly connected at their origins, nor during the first three decades of their emergence. Efforts to analyze the relationship between human rights law and indicators of states' compliance did not begin in earnest until the turn of the 21st Century. Since then, two main approaches have come to dominate a quickly expanding body of research. One we term

¹ These counts are themselves imperfect indicators. For the number of protected rights, see Green (2001). For numbers of treaties and instruments, see the UN Office of the High Commission of Human Rights (<http://www.ohchr.org/EN/ProfessionalInterest/Pages/UniversalHumanRightsInstruments.aspx>, accessed August 22, 2016). Elliot (2011) places the number of total human rights instruments at 779, though his definition is looser, and extends between the years 1863 and 2003.

² In addition to the indicators produced by Freedom House, the Political Terror Scale, and the Cingranelli and Richards team, new, and over 300 expert-coded democracy, rule of law, and human rights indicators that were recently published as part of the Varieties of Democracy Project.

the factualist approach,³ which studies the relationship between law and human rights violations using indicators that are accepted as accurate representations of social facts. The other is the constructivist approach, which does not treat indicators as facts but as socially produced technologies of knowledge and power. Though these perspectives diverge radically, they agree on a central notion: that international human rights law has contributed very little to social progress. As we will explain in the first two sections, these approaches have reached their apotheosis, converging on a number of irresolvable challenges.

In this article, we synthesize an emerging constitutive approach to law and indicators that means to address shortcomings in the literature by connecting theories of human rights with theories of measurement. The constitutive approach argues that human rights law possesses a politically motivated normative dimension that fundamentally changes the way humans process and understand data. The law did not cause indicators, but it has changed them over time. The constitutive approach addresses constructivists' central criticism--that the production of statistical indicators is not a neutral or independent social process. In so doing, it also challenges the conclusions from the factualist approach by questioning the assumed validity of indicators. Instead of rejecting indicators outright, as the critical constructivists often do, the constitutive approach points to a set of research design and statistical tools that allow for the linking of theory and measurement.

The constitutive approach encourages scholars to treat measurement itself as an object of theorizing and inquiry. Fortunately, we now stand on the verge of a revolution in human rights theory and measurement, whereby social scientists can more accurately estimate the nexus between law, indicators, and state behaviors. Researchers are now able to use new computationally intensive but theoretically appropriate tools for understanding the flaws and biases of the information produced to understand human rights. When

³ We adopt the term factualism to indicate an excessive faith in the ability of numbers to correspond to factual truth—and to construct statistical models with little worry about the violation of modeling assumptions. Others have referred to this as a “mechanical” approach to statistics (Achen 2005). In the language of research design, the factualist approach rests on the assumption that data are perfect empirical representations of the theoretical concept or, if some uncertainty is acknowledged, is dismissed as unrelated to the relationship modeled using the data.

applying these new theoretically motivated tools, it becomes clear that not only is there a relationship between law and data collection, but also, once this relationship is taken into account, that law has promoted improvements in the human condition. This is the promise of the constitutive approach.

II. A Genealogy of Human Rights Indicators

In 1976, the year that the two main international human rights covenants took legal effect, the US State Department already had an institutional architecture in place for reporting on human rights practices and democracy in other countries. So did Amnesty International and Freedom House. However, none of these organizations had yet initiated standardized yearly reporting in order to monitor compliance with human rights laws. Embittered by Vietnam and the coup in Chile, and intent on linking foreign aid disbursements to improvements in rights protections, the US Congress established the State Department's Bureau of Human Rights and Humanitarian Affairs in 1975. The Bureau would eventually develop into a fully functional human rights reporting organization (Cingranelli & Pasquarello 1985, De Neufville 1986).⁴

Amnesty International (AI) developed its own reporting methods a decade prior, in response to a *lack* of effective legal treaties (Clark 2001, Grant 2011). If states would not take concrete measures to enforce human rights, AI argued, it would begin monitoring compliance on its own. AI published its first annual report in 1962, and performed a successful monitoring mission in authoritarian Greece in 1968. The organization followed up in 1973--the same year Donald Fraser initiated first-of-a-kind Congressional Hearings on human rights in US foreign policy--by starting a very influential campaign against torture, complete with more extensive reporting that set a standard for human rights documentation (Rodley 2011). This campaign set in motion a process that ended in the creation of the Convention Against Torture (CAT) over a decade later (UN General Assembly 1984). AI's expansion and notoriety coincided with

⁴ This did not happen immediately. Secretary of State Henry Kissinger bristled at the change, and defiantly assigned an obscure Ambassador to Micronesia, James Wilson, to head the Bureau. However, the Carter Administration diligently developed these reports, and despite early resistance from President Reagan, the reports continued to get more reliable in the 1980s (Keys 2010).

the release of Freedom House's first "comparative survey of freedom" report in 1972. An anti-Communist outfit funded in part by the State Department, Freedom House had scored countries according to freedom "debits" and "credits" since 1955, but in 1972 switched to categorical seven-point scaled rankings of civil and political rights (Bradley 2015, Giannone 2010, Satterthwaite 2016). All of these early human rights reports—which relied on foreign service officer documentation, witness testimony, and expert rankings—were inspired in some way by the concept of human rights, but none purported to measure directly compliance with legal treaties. The law was too underdeveloped by the time the reports had started to accomplish this goal.

Scholarship using data on human rights violations also did not focus much on the question of compliance or implementation of international law. One exception is Matthew Lippman's 1979 article in the first volume of *Human Rights Quarterly*, which notes that 61 countries use torture and that "it is futile to attempt to protect individuals from torture through treaties and legal instruments..." The article counted binary data from Amnesty International on whether torture is used at all in a country. The extent of the methodology was thus to report any instance of torture as a failure measured against an ideal of zero reported incidents. Other more sophisticated studies from this early period focused on whether military regimes were more likely to suspend or violently suppress civil and political rights (Claude 1976, Henderson 1982, Hibbs Jr. 1973, McKinlay & Cohan 1976), or whether the US government's efforts to slow foreign aid to rights-violating countries worked (Cingranelli & Pasquarello 1985, Poe 1990, Rubin & Newberg 1980, Schoultz 1981, Stohl et al. 1984). As of 1984, the year that the Convention Against Torture (CAT) was adopted, the two strands of research using human rights indicators were comparative studies of repressive regimes and analyses of US foreign policy.

Some seminal attempts to measure human rights, like Charles Humana's *World Human Rights Guide*, came and went during this early phase (Humana 1983).⁵ In the process, Freedom House's civil and political rights indices gained the most cachet by the early-1980s. This was the data of first resort for two main reasons. One is that Freedom House produced two simple ordered categorical scores for each country, which allowed

⁵ See also (Rummel 1976, Taylor & Jodice 1983).

for longitudinal comparisons back to the early 1970s (Bollen 1986). The second reason is that it employed a checklist developed and published by a social scientist, Raymond Gastil (1980). This checklist provided some transparency. Perhaps in response to this popularity, many scholars voiced concern that Gastil's scorings, which he undertook individually, were not reliable because they could not be replicated (Scoble & Weisberg 1981). Other researchers showed that, because they were so highly correlated, the civil and political rights scales were largely redundant, and that the coding was biased toward Western understandings of liberal democracy (Banks 1986, Bollen 1986).

The years 1984-1986 marked a turning point for the use of human rights indicators. Two things happened. First, scholars began to devote a significant amount of attention to the subject of measurement itself. A special issue of *Human Rights Quarterly* in 1986 focused on the statistical methods and human rights, as did edited books in 1986 and 1988 (Cingranelli 1988, Stohl & Lopez 1986). In the 1986 *HRQ* symposium, two points of view started to solidify. One held that the study of human rights was plagued by measurement issues, but that these could be addressed with greater methodological rigor and theoretical attention to the process by which the information was generated (Stohl et al. 1986). The other held that human rights scholarship was blindly following the dictum that "if data exist, they will be used" (Goldstein 1986). From this perspective, the tendency to fetishize numbers means preferring bad data to no data, risks a loss of interpretive nuance, and creates a fundamental alteration to the production of knowledge. These themes would continue to resonate in future work, and remain present today.

A second big change that came about in the mid-1980s was the exit of Freedom House as the most relied-upon measure of human rights. Freedom House was repeatedly challenged on methodological grounds, and by 1989, Raymond Gastil was replaced by a panel of coders. He admitted during this time that he had sometimes coded countries based on unsystematic impressions (Barsh 1993, Spierer 1990). Interestingly, toward the end of the Cold War, the Freedom House indices shifted from the human rights measures of choice to the democracy measures of choice. In so doing, it moved from use primarily as a dependent variable to an independent variable in statistical models of human rights. Though the FH indices have also been challenged as a poor indicator of democracy (Munck & Verkuilen 2002), they continue to be favored among government officials, and

policy-oriented scholars who publish in outlets like *the Journal of Democracy* or *Foreign Affairs* (Diamond 2011, Giannone 2010).

As Freedom House left the human rights scene, two other global databases arose to fill the gap. One is called the Political Terror Scale (PTS). The first iteration of this project was associated with the researchers at Purdue University.⁶ Under the directorship of Mark Gibney since 1984, PTS adapted a coding scheme from the Freedom House checklist, and applied it to information systematically captured in *The Country Reports on Human Rights Practices* published annually by the US State Department and *The State of the World's Human Rights* report published annually by Amnesty International (Wood & Gibney 2010). The resulting 5-point categorical scales are based on the content of each set of reports. These two scales measure the extent to which every state in the world commits abuses to physical integrity rights, dating back to 1976.⁷

The other project, the Cingranelli-Richards Human Rights Data Project (CIRI), came out of David Cingranelli's early work on US foreign policy (Cingranelli & Pasquarello 1985, Cingranelli 1988), but it did not become publicly available until 1999 (Cingranelli & Richards 1999a). These variables, ranging from 1981-2011, are also drawn from State Department and AI reports, but CIRI introduced a new scheme to disaggregate physical integrity violations combining information from both sets of reports. The set of variables contains four categorical scales (each 0-2) representing the extremity of political imprisonment, torture, extrajudicial killing, and disappearance in a country. Combined, these sub-scales produced an additive 9-point categorical scale: Physical Integrity Index (PHYSINT).⁸ This scale, somewhat confusingly labeled an index, is the summation of four, 3-point categorical scales.⁹

⁶ These include Michael Stohl, George Lopez, and David Carleton.

⁷ Fewer countries were reported on by the US State Department before 1981.

⁸ The project also used information in the reports to code an expanded set of categorical scales (each 0-2) representing the extremity of assembly and association, foreign movement, domestic movement, speech, electoral self-determination, religion, these sub-scales are also used to produce an additive 15-point categorical scale: Empowerment rights index. The CIRI project contains two versions of some of these variables and therefore two versions of the additive empowerment scale. Finally, the project also coded several categorical scales measuring women's rights.

⁹ Both an "index" and a "scale" are models that combine multiple pieces of data into a composite measure. Generally though, indices and scales combine different forms of

PTS and CIRI would come to dominate the statistical study of human rights. Since 1982, PTS alone has been used in just over 500 scholarly publications.¹⁰ However, the conquest of PTS and CIRI did not happen immediately. It followed a period during which categorization and statistical analysis of human rights and repression went mainstream. From 1987-1995, authors sought to establish patterns linking structural determinants to rights violations, using many different combinations of indicators. Park (1987) showed that Islamic states were less likely to protect civil and social rights. Mitchell and McCormick (1988) found that developed states were more likely to protect physical integrity. Harff and Gurr collected data on genocides, politicides, and minorities at risk, and argued that fewer of these are found in Europe and Latin America (Gurr & Scarritt 1989, Harff & Gurr 1988).

New statistical discoveries dovetailed with the “end of history” mentality in the last throes of the Cold War (Fukuyama 2006). In 1991, amidst the dissolution of the Soviet empire, UNDP released a report ranking all countries according to their adherence to the Universal Declaration of Human Rights. The critical tone of this report sent lesser-developed and former-Communist countries into an uproar, but resonated with human

information. Indices combine ratio or interval level data whereas scales combine categorical data that is ordered or unordered (nominal). For example the Dow Jones Industrial average is an index that combines the stock prices (ratio level data) for 30 publicly traded companies. A scale tends to combine categorical responses such as “strongly agree,” “moderately agree,” “moderately disagree,” and “strongly disagree”. In some cases these categories map on directly to interval level numerical information. For example, take “hot” “warm” and “cold.” This is an English language-based measure of the concept temperature with three constituent parts. If one were to substitute 1, 2, and 3 for “hot,” “warm,” and “cold,” the following statement would make perfect sense: “Boiled water is very 1, and ice cubes are very 3.” In this respect, both the alphabetical and numerical indicators are exactly the same, containing no loss of meaning. However, interval or ratio expressions are often preferred for their precision. If one wants to know the temperature outside, one most likely consults a thermometer, which ranges from -50 to 150 degrees Fahrenheit. Thermometers do not possess an inherent advantage over categorical indicators, but they capture more exacting variations in temperature than the English language allows. It would be quite difficult to memorize and apply 200 different words describing the temperature outside! The greater precision that thermometers provide us allows for comparison of temperatures across locations, and the ability to know with certainty how bad a fever is. Generally, in the human rights field, we only have categorical information with which to combine into variables so the term scale is preferred here.

¹⁰ See “Bibliography,” www.politicalterroryscale.org.

rights statistical research (Barsh 1993). For example, based on the most comprehensive longitudinal categorical data from PTS, and using the most sophisticated statistical techniques available at the time, one prominent study showed that democracy was definitively better than Communist dictatorship (Poe & Tate 1994). Government in the West was superior, and the data verified it. With these developments, human rights indicators went from a specialist's field to a part of the post-Cold War zeitgeist.

Though the mainstreaming of human rights indicators coincided with the “decade of international law,” scholarship still did not draw a direct connection between international law and indicators of rights protection. Studies in the late 1990s basically focused on the multivariate structural correlates of PTS and CIRI scores, combined with a theory concerning leaders' rational decisions to repress (Mason & Krane 1989). The main findings were that democracies outperform other regimes (Armstrong & Davenport 2004, Cingraneli & Richards 1999b, Davenport 1996); that transitional or middling democracies are actually more violent than stable democracies and autocracies (Fein 1995, Regan & Henderson 2002); that regimes respond to dissent and other threats with repressive violence (Carey 2009, Davenport 1995); and that civil war is consistently associated with more violations to physical integrity (Poe et al. 1999). Combined, these factors formed a “standard model” of repression (Keith 2012). However, this model included nothing about states' obligations to international human rights law. By the turn of the century, human rights lawyers had managed to push through a number of human rights instruments, and scholars of social movements were high on the potential for legal activism to address a number of rights (Keck & Sikkink 1998). However, statistical human rights research to this point engaged sparingly with law.¹¹ Comparative human rights studies possessed an advanced and evidence-based sense of what structures threaten a narrow set of “survival rights,” but repression studies remained disconnected from the broader international legal context (Donnelly & Howard 1988).

III. Two Approaches to Law and Indicators

¹¹ One example of sparing engagement is the CIRI codebook, which mentions that physical integrity rights are protected under human rights law. But functionally, this means little in terms of the way the data are collected.

In some respects, the disconnect between comparative and legal research is to be expected. Political scientists are not lawyers. However, by 2000, there was a growing sense that political science should more deeply engage law in international and comparative politics. In that year, *International Organization* printed a symposium devoted to the concept of legalization in international relations. Others lamented the great scholarly expanse between international law and political science (Raustalia & Slaughter 2002). The effort to combine legal and comparative scholarship using indicators began in earnest after the initiation of the War on Terror.¹² In 2002, Oona Hathaway (2002) published an article demonstrating that countries which ratify the ICCPR are, counter-intuitively, more likely to commit rights violations.¹³ This was followed by a number of “compliance” studies that test the effects of international law by correlating treaty ratifications with PTS and CIRI scores (Conrad 2012, Conrad & Ritter 2013, Hafner-Burton & Tsutsui 2005, 2007, Hill 2010; Lupu 2013a,b, 2015; Neumayer 2005, Powell & Staton 2009, Simmons 2009, Simmons & Danner 2010). One study claimed human rights legalization and measured rights practices are “radically decoupling” (Hafner-Burton & Tsutsui 2005), while others find that treaty ratifications are associated with declining repressive violence in countries with an active NGO sector, or that are transitioning to democracy (Powell & Staton 2009, Simmons 2009). Still others argue that, on the basis of this correlational evidence, autocratic leaders ramp up abuses in the wake of treaty ratification, or “use” human rights treaties to signal their lack of willingness to leave office (Hollyer & Rosendorff 2012, Vreeland 2008).

The style of these compliance studies—where state treaty ratification is treated as a predictor, and changes in that state’s average physical integrity scores are treated as an outcome—embody what we call the factualist approach to international law and indicators. The factualist approach usually starts by deducing a causal theory relating an independent variable (causal input) to a dependent variable (social outcome). The aim for the scholar studying human rights law using the factualist approach is normally to evaluate whether, or under what conditions, multilateral treaties --- with monitoring

¹² One clear exception is Camp Keith (1999).

¹³ Hathaway used her own five-point scale of torture, which looks very similar to the CIRI torture index.

bodies but little direct enforcement --- can improve the recognition of rights at the country-year level. Thus, the independent variable is typically a legal treaty like the ICCPR or CAT, and the dependent variable is the categorical level of human rights abuse from the PTS or CIRI projects.

Most factualist models make two assumptions. The first is that legal treaties and repression variables are exogenously related (or can be modeled as such), meaning that they belong to distinct social processes. Researchers test theory by studying the correlation between two data series normally retrieved from already-existing projects. Independent indicators of law are usually binary categories —1 for treaty ratification, 0 for no ratification—and indicators of human rights abuse are the PTS or CIRI categorical scales. Because these data are available across time in many countries, it is possible to estimate a value representing the average categorical physical integrity scores in all countries and years in which laws have been ratified, and compare those averages to all country-years in which laws have not been ratified. A second assumption is that these data series are measured without error or bias. That is, there is no omitted variable that causes change in these variables either directly or through the process by which they are measured. Implicit to this approach is trust that the procedures used to generate the information and measure these concepts are not systematically related to the development of human rights law or the actualization of state human rights practices.

The factualist approach to human rights law uses statistical models as an analytic tool but takes for granted the process by which the data are generated. Because the approach takes measurement for granted, it has the advantage of being relatively simple to implement, and its contributions are powerful because they seem both elegant and systematic. It is difficult to argue with averages when the precision of the underlying data is presumed to be perfect. For example, Eric Posner contends in 2014's *The Twilight of Human Rights Law* that “there is little evidence that human rights treaties, on the whole, have improved the well-being of people, or even resulted in respect for the rights in those treaties” (p 7). As evidence, Posner shows charts indicating that increasing numbers of treaty ratifications over time are not accompanied by concomitant upward-sloping averages in CIRI scores. A similar contribution comes from Rod Abouharb, Mikhail Fillipov, and David Cingranelli. In a 2015 volume, they find with regressions that

ratification of the ICESCR is correlated with higher scores on the Physical Quality of Life Index (PQLI). However, “ratifying more international covenants worsens ESR [economic and social rights]” because too many laws means too many demands on leaders, who experience “agency loss” (p 43). The statistical models used in these studies provide insight into why rational leaders seemingly fail to live up to the expectations embedded in the human rights treaties ratified by their countries.

In many ways, these two studies represent the apogee of the factualist approach. Statistical work has left us with a sea of correlations, negative and positive, between ratifications, domestic institutions, and human rights indicators (Hafner-Burton 2014). But no matter how one cuts it, law’s impact is on average marginal, highly conditional, and causally underwhelming. Thus, the common interpretation is that law has not contributed significantly to progress. This is not the whole story. It is a version of reality heavily influenced by the way indicators are created. There is little reference to or theory about the production process that originally generates them.

The factualist account is also narrow and process-weak. It is narrow because compliance is operationalized as changes in aggregate state practices that follow in short order after ratification. However, this definition of compliance replaces a regulatory concept of rule-following with a consequentialist concept of effectiveness.¹⁴ Official actors who in principle endorse the use of numerical indicators contend that they should not only capture outcomes, but also measure legal processes that intercede between treaties and aggregate state practices (Landman 2004, Landman & Carvalho 2009). In a 2008 report, the UN’s Office of High Commissioner of Human Rights (OHCHR) embraced the use of indicators, but suggested a three-part framework for measuring “structure, process, and outcomes.” Factualist studies are weak on process because most do not collect or analyze data on the laws and behaviors that connect specific treaty provisions to domestic legal practices. One reason is that these data would require painstaking collection efforts, and might not lend themselves to cross-sectional time-

¹⁴ “Assessing compliance is restricted to the description of the discrepancy between the (legal) text of the regulation and the actions and behaviors of its addressees. Perfect compliance, imperfect implementation and zero effectiveness therefore are not necessarily mutually exclusive” (Neyer & Wolf 2005). See also Finnemore & Toope (2001) and Howse & Teitel (2010).

series analysis.

Another reason the factualist approach is limited is that categorical indicators might be fundamentally unable to capture the true process of compliance. This position is central to the constructivist approach to law and indicators. Constructivists hold that indicators are not a neutral or objective way of measuring reality. Instead, they *produce* a particular version of reality. They are a form of power-knowledge that reduces dynamic legal processes, political contestation, heterogeneous local contexts, and cultural symbols to static numbers (Davis et al. 2015).¹⁵ Further, the practice of making indicators shifts legal-political matters to the field of accountancy. It therefore reinforces a rationalist “audit culture” that assigns worth only to those interventions that have a quantifiable impact (Rosga & Satterthwaite 2009). At a maximum, this is abusive because indicators are produced by the powerful to legitimize hegemony, albeit behind the veil of objectivity. A case in point are the Freedom House indices (Bradley 2015, Giannone 2010). At a minimum, law is given short shrift because its normativity, or its ability to set standards and set in motion influential political interactions, is almost completely dismissed by the factualist approach. As a technology, numbers simply cannot represent norms like words and symbols can. Categorical data is therefore too crude to capture the conceptual richness of law, human rights, and the relationship between them.

Critical constructivist accounts of human rights law are not the first to lament that statistics are a language of control that does violence to variation in local context. After all, we are reminded that the word “statistics” itself derives from the German *Stat-istik*, which was used to describe a practice of collecting and rationalizing information for the purpose of state-building. These practices were damaging, harming nature and killing people (Scott 1998). Yet the factualist approach to data often promotes the use of statistics precisely because it allows one to see through the noise of local context (Hafner-Burton & Ron 2009). However, for constructivists, adopting this quantitative “language of distance” when considering human rights will cause one to miss diverse factors that require political judgment (Rosga & Satterthwaite 2009). The founding creators of human rights indicators understood this,¹⁶ but the message got lost as social scientists were

¹⁵ For power-knowledge, see Foucault (1980).

¹⁶ “Similar abuses may manifest themselves differently in different political contexts.

trained to study law and politics by downloading data sets and crunching the numbers without reference to the underlying processes those data actually represent (Adcock & Collier, David 2001, Schrodtt 2014).

In addition to loss of context, critical constructivists identify three concerns regarding the use of indicators in the study of human rights law. The first is that preference will be given to questions that lend themselves to quantifiable answers. Research on law will become data-driven especially when the data are easy to collect. One may observe this phenomenon regarding economic and social rights (ESR). The literature on human rights law contains a tremendous amount of studies of physical integrity rights. Well-designed and implemented studies of civil and political rights are also available, though less plentiful (Davenport 1996, 2007). But studies of economic and social rights are very few. Contrary to many accounts, this is not because human rights scholars from the West were traditionally uninterested in these rights. In fact, many scholars in the early period of research wrote thoughtfully on ESR (Donnelly & Howard 1988, Howard 1983). Others thought that valid indicators of ESR already existed, meaning that more attention should go to collecting data on other sets of rights.¹⁷ However, recent research has shown that conclusion to be exactly wrong. Legal protections like the right to health or the right to education, which are subject to “progressive realization” under international law, are tremendously difficult to measure (Kalantry et al. 2010). Doing so requires the creation of benchmarks, careful legal thinking, and complex frameworking. Because this does not translate well into cross-national and quantitative indicators—for example, percentage of primary school enrollment is a poor indicator of compliance with international law---scholars over time stopped trying to collect much data or analyze of these kinds of rights (Rosga & Satterthwaite 2009).¹⁸

And, thus, in short, the statistical analysis of human rights must always be informed by political analysis” (Stohl et al. 1986, p 598).

¹⁷ “The area of social and economic rights lends itself to much easier definition and operationalization than the political, civil, and personal security rights area. Thus, there is general agreement on the definition of terms such as infant mortality, life expectancy, and caloric intake” (Goldstein 1986, p 611) .

¹⁸ Banks (1986) gave reasons for omitting ESR variables from his analyses in 1986, as did Andrew McNitt (1988).

A second problem is Goodhart's Law, or the "tendency of a measure to become a target" (Rosga & Satterthwaite 2009). When human subjects are measured and held accountable on the basis of indicators, they adapt. Part of this adaptation is to improve on the measured item, without necessarily changing the structures that affect performance. If it is policy to allot funds to schools on the basis of test scores, then the schools will "teach to the test" to improve the scores. The indicator being used to hold schools accountable ceases to be a valid measure of overall education because it is now systematically related to test preparation instead of educational attainment. A similar process occurred in the advocacy and study of physical integrity rights. If a government knows its repressive tactics are being scrutinized, it will shift in order to hide its abuses. DeMars (2005) argues that the innovation of disappearances in Argentina and Chile during the 1970s was actually a response to human rights reporting. Because Amnesty International criticized these countries so heavily for torture, the regimes made their enemies disappear, rendering torture undetectable.¹⁹

If Goodhart's Law holds, and we think that it does, it means that indicators themselves change practice. Some quantitative human rights researchers have long recognized this possibility, but up until recently the standard approach of the factualist data user was to mostly ignore it (Barsh 1993, Goldstein 1986). A third and related problem is indicator endogeneity. This problem refers to the possibility that the construction of indicators is itself a function of the law, and indicators change as the law changes. Clark and Sikkink argue the "good news about increased human rights information [is] bad news for human rights measures" (Clark & Sikkink 2013). Because digital media and the diffusion of transnational information networks both expanded since the origin of human rights research, there is a distinct possibility that researchers producing indicators see more evidence of abuse than they did 30 years ago. More evidence means worse scores, and worse scores over time means deceptively little change in estimates of average human rights practices. Because international legal bodies engage in monitoring and information-production (SIMMONS 2009), law and the legal processes that accompany it, are indirectly implicated in the production of indicators.

¹⁹ See also Ron (2000) and Berman and Clark (1982). For shifts in tactics of torture that make it more difficult to investigate, see Rejali (2007).

The constructivist approach alerts us to these possibilities, but it falls short on solutions. Newer work in the critical constructivist vein tends to strike a chord of despair. Stephen Hopgood's *The Endtimes of Human Rights* (2013) is a powerful critique of the culture of law-giving, standardization, and imperious universality that pervades the human rights legal regime. Though he resists wholesale the econometric method, Hopgood shares with the factualist approach the conviction that human rights law contributes little to progress. The book ends with a prediction that international criminal law will soon crumble. Another critical volume devoted to indicators ends not with a way forward, but with an open-ended question: "Why would anyone want to use an indicator that put itself radically in doubt?" (Davis et al. 2015).

What we are left with, then, are two approaches that are pessimistic about the prospects for law, but which diverge significantly about what to do and how to evaluate the efficacy of rights-protecting institutions. The factualist approach continues to use human rights indicators to produce findings, but is only beginning to confront their shortcomings. The constructivist approach seems content to persistently sound an alarm about indicators based on the following problems: (1) inattention to context, (2) ignorance of rights other than protections of physical integrity (3) the tendency of measured subjects to adapt to measurement and (4) the endogenous relationship between law and the production of indicators. Next, we outline a third constitutive approach to human rights law and indicators. With this approach, we attempt to address these four problems, while also demonstrating that human rights law is associated with measurable improvements in the world.

IV. A Constitutive Approach to Information, Law, and Indicators

Law does not "independently" cause the production of indicators. However, law directly influences how observers and academics use information to generate, develop, and change indicators. One pathway for the direct influence of human rights law is through the reporting requirements for UN treaties (Creamer & Simmons 2015). Once states ratify a human rights treaty, they must develop new domestic institutions or make use of existing ones to monitor, document, and report on the implementation of the provisions of the treaty (Creamer & Simmons 2015). All UN human rights treaties

contain such provisions. Information gathering and reporting -- measurement -- is therefore an important type of compliance behavior that varies between states and over time. International and domestic monitors are also able to look harder and in more places for abuse because of an increasingly dense network of civil society organizations (Greenhill 2010, Keck & Sikkink 1998, Murdie & Bhasin 2011). Sometimes these civil society organizations connect and work with the state institutions charged with collecting information about compliance with domestic and international law. What are we to make of the complex relationship between law, measurement, and shifting human rights practices?

We could follow the factualist approach and continue to use human rights indicators to produce statistical findings while downplaying the biases of the data (Hafner-Burton 2014). Or, we could follow the suggestions of constructivists and, cognizant of the failings inherent in the data, largely reject the use of indicators entirely. Neither of these pathways offers a way forward.

Fortunately, a third way is now coming into form. The constitutive approach suggests that human rights law possesses a normative political dimension that fundamentally changes the way human beings process and understand information (Dancy & Fariss 2017). The constitutive approach is built on two theoretical assumptions. The first assumption is that law, human rights compliance, and the production of indicators are not independent processes. They are endogenous and mutually constitutive, in part because binding human rights law and human rights information-gathering emerged as part of the same transnational campaigns (Brysk 1994, Keck & Sikkink 1998). Information is never neutral, apolitical, or objectively separate from legal mobilization. This idea compels one not only to theorize about the relationship between law and human rights compliance as concepts of theoretical interest but also to theorize about the monitoring and observational processes that have developed over time to produce information and indicators.

In reference to the science of indicators, the constitutive approach suggests that context in which information is produced affects the translation of the theoretical concept into an operational protocol. Translation error is a mismatch between the theoretical concept being represented and the proposed pieces of information chosen by the analyst

for inclusion in the construction of that concept's indicator. This occurs when the operational protocol does not match the theory of the concept. If one were measuring judicial independence, it would be a translation error to propose including information on popularity of Supreme Court decisions. Though popularity might make a court more effective, it does not mean the court is more independent, which applies to its judicial autonomy. Translational errors are often more subtle and difficult to detect than measurement errors for three reasons. First, legal concepts that scholars operationalize are complex and multifaceted. Civil and political rights like freedom of expression, or economic and social rights like the right to education, look very different in case law than as indicators. Second, monitoring agencies or institutions creating the information used to construct indicators are often not the same entities that are building theoretical concepts. Researchers at AI and the US State Department are often surprised to find how extensively academics use their reports for data analysis. And third, as agencies change strategies over time, they also change the operational protocols they use for monitoring. If the analyst accepts the possibility of translation error, she will need to assess the original operational protocols used to collect existing information, alongside the theoretical concept it purports to represent.

The second assumption of the constituent approach is that, even once translation errors have been addressed, measurement error exists because both systematic processes and random chance affect the collection of the information used to create data. Measurement errors occur when data collection procedures used to collect information and then generate indicators miss the true value of a concept. The true value of the theoretical concept is what we call the "latent" trait because, though we believe it exists, it is difficult if not impossible to objectively and precisely observe. This holds for both quantitative and qualitative data. For example, only one true value of yearly police killings in the US exists. Likewise, there is only one value of police accountability for those killings. Estimating the value of police killings requires a sophisticated approach that accounts for measurement error (Ball 2016). It is difficult to ascertain the latent value of police killings with certainty because killings go unreported, data on lethal violence is not collected in a uniform way by local law enforcement, and the FBI makes mistakes when aggregating information from every locality into a database. Police accountability

is an even more difficult concept to measure. Not only does it require counting events like trials of police officers, it also requires translating those events into a coherent and justifiable concept of accountability.

Most indicators of human rights tend to base categorization on *available* information about a specific rights practice. If the availability of information relevant to the operational protocol is missing, either systematically or randomly, then measurement errors will be present in the resulting indicators. Measurement errors are often addressed by collecting new data with special attention paid to why information might be missing systematically or at random. Valid comparisons require the careful and tedious application of the theoretically appropriate operational definitions concerning human rights abuses from incident to incident. The amount of work involved is often mind-boggling, especially if data must be gathered from scores of newspapers, archives, witness testimonies, or police files (Goldstein 1986). Even with new theoretically appropriate operational protocols, the collection of new data, however studious the data collector may be, will still contain at least some random error and possibly systematic error as well.

The sheer weight of these concerns inspires some scholars to counsel against systematic comparison in favor of practical judgment based on deep knowledge of cases (Zunino 2011). However, one may accept the constitutive model's assumptions and still attempt valid comparison. One pragmatic approach is to use models that can estimate and correct for translation and measurement errors using existing data. A measurement model (e.g., a latent variable model) is a statistical tool that brings together diverse sources of data (both quantitative and qualitative) and links them together in a theoretically informative way. This is accomplished within the model either (1) by assessing translation errors between the theoretical concept and the operational definition, or (2) by addressing measurement errors that occur when the operational protocol is imperfectly employed to generate or gather the data. In both cases, these discrepancies can be adjusted for within the structure of the model and then evaluated empirically.

One example of a productive measurement model is Fariss's (2014) Human Rights Protection Scores Project. Fariss addresses a translational issue by arguing that the operational definition used to generate human rights data based on information from

human rights reports has changed over time even though existing datasets like PTS and CIRI treat all cases as though they are directly comparable. Though the operational procedure of the CIRI and PTS projects reliably capture information from the human rights documents, the operational production of the reports themselves changed over time. The production of information contained in these reports was related to changes in what monitors consider human rights abuses and changes in where and how hard the monitors looked for abuses. The model and the new insights it produced are based on an identifiable translation error between the concept of human rights and the operational protocol employed by AI and the US State Department. The model then, without need of new data collection, is modified to account for the possibility that the standard of accountability has changed over time. This work, which goes to great pains to address translation and measurement errors, finds that progress has occurred in specific issue areas like physical integrity rights, and likely some other areas as well (Figure 1). Moreover, improvements in physical integrity over time are correlated with the increasing density of international laws being ratified and implemented by states (Fariss, forthcoming).

Other measurement projects with theoretical insights about the way the information is produced provide a counter point to both the factualist and constructivist approaches (Crabtree & Fariss 2015). New data projects adopting a constitutive approach to understand the relationship between new and existing sets of indicators and theoretical concepts are in development. New human rights indicators were recently published as part of the Varieties of Democracy Project (Coppedge et al. 2014). The expert-coded data includes several hundred new indicators of human rights, democracy, the rule of law, and other institutional features from 1900-2015. The measurement models used to construct these latent variables make use of multiple country expert coders who answer categorical questions for each country-year unit. This project includes many new human rights variables, two of which are physical integrity variables: (1) freedom from political killing, and (2) freedom from torture. Like Fariss (2014), the V-DEM data show a stark improvement over time (Figure 2). Though the V-DEM measurement model attempts to address the disagreement between coders, bias might still persist if the expert-coders are using the same historical source material or share common

biases. But these pitfalls actually represents the value of this measurement approach because the exploration of these potential biases and how they relate to the biases from other human rights data are important areas of future research.

Measurement models are now being deployed to create unified measures of other concepts like judicial efficacy (Figure 3) and women's rights protections (Hill & Inglett 2015, Linzer & Staton 2015). Thus far, these projects are capable of generating a fuller account of historical trends based on information captured by many data sources. In this sense, measurement modelers are moving beyond the presentism of many singular studies, which tend only to capture temporally limited amounts of information, with a great deal of bias and error. In this sense, the new measurement modelers are escaping the narrow islands of empirical analysis produced by factualist accounts. But they are also moving beyond constructivist discontents, who see a complex world composed of human right violations with ever-denser legal regulations, made even more complex by the hard-to-manage ecology of indicators diffusing in fits and starts across a chaotic global information network. Proponents of the constitutive model take this complexity seriously, and they are beginning to cut through it.

V. Conclusion

Fifty years ago, the world had very few human rights laws, very little information on human rights violations, and very few indicators of those violations. Today, the situation could not be more different. We are awash in all three. The factualist response has been to measure what one can, however imperfectly, and establish as many correlations between “independent” variables as possible. The constructivist approach is to remind us that these processes are not independent, and that our efforts as a science of law and human rights is fallible. The constitutive approach accepts the constructivist critique and offers a way forward.

The constitutive approach to data analysis supports macro-historical work that finds evidence of legal-normative progress (Sikkink 2011), lessening cruelty (Pinker 2011), and declining war over time (Goldstein 2012). However, its great promise is that it also provides for micro-level comparative analyses of singular cases of social decay amid this larger backdrop of positive change. This approach can thus put to rest long-standing

arguments over the relative utility of indicators, while also enabling a more concrete science of law and human rights (Schnakenberg and Fariss 2014).

But scholars are only at the tip of the iceberg. Much work remains for adherents to the constitutive approach. Measurement theorists should work with treaty bodies to establish more reliable indicators of economic and social rights like the right to education and freedom from racial discrimination. Furthermore, while macro-level relationships between expanding international legal regulations and improved practices is becoming evident, more research needs to be devoted to the mechanisms linking those processes together. For example, one research program demonstrates that state commitment to international rights law is correlated with greater activism in those states (Simmons 2009), as well as greater efforts at domestic litigation and transitional justice (Dancy & Sikkink 2012, Dancy & Michel 2016, Dancy & Montal 2017). However, the same measurement techniques currently being applied to indicators of human rights *violations* need to be applied to these mechanisms of *remedy*. Studies of what is to be done need to keep pace with studies of what is wrong. Only if scholars move toward more comprehensive and theoretically sound indicators of mechanisms, in addition to indicators of causes and outcomes, can a science of law and humans be advanced.

References

- Achen CH. 2005. Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong. *Confl. Manag. Peace Sci.* 22(4):327–39
- Adcock R, Collier, David. 2001. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *Am. Polit. Sci. Rev.* 95(3):529–546
- Armstrong D, Davenport C. 2004. Democracy and the Violation of Human Rights: A Statistical Analysis from 1976 to 1996. *Am. J. Polit. Sci.* 48(3):538–54
- Ball P. 2016. Violence in Blue
- Banks DL. 1986. Analysis of Human Rights Data over Time, The. *Hum. Rights Q.* 8:654

- Barsh RL. 1993. Measuring human rights: Problems of methodology and purpose. *Hum. Rights Q.* 15(1):87–121
- Benenson P. 1961. The Forgotten Prisoners. *Lond. Obs.*, May 28
- Berman MR, Clark RS. 1982. State Terrorism: Disappearances. *Rutgers Law J.* 13(3):531–78
- Bollen KA. 1986. Political rights and political liberties in nations: An evaluation of human rights measures, 1950 to 1984. *Hum. Rights Q.* 8(4):567–591
- Bradley CG. 2015. International Organizations and the Production of Indicators: The Case of Freedom House. In *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*, eds. SE Merry, KE Davis, B Kingsbury. Cambridge: Cambridge University Press
- Brysk A. 1994. *The Politics of Human Rights in Argentina: Protest, Change, and Democratization*. Stanford: Stanford University Press
- Camp Keith L. 1999. The United Nations International Covenant on Civil and Political Rights: Does it Make a Difference in Human Rights Behavior? *J. Peace Res.* 36(1):95–118
- Carey SC. 2009. *Protest, Repression and Political Regimes: An Empirical Analysis of Latin America and Sub-Saharan Africa*. London and New York: Routledge
- Cingranelli DL, ed. 1988. *Human Rights: Theory and Measurement*. St. Martin's Press
- Cingranelli DL, Pasquarello TE. 1985. Human Rights Practices and the Distribution of U.S. Foreign Aid to Latin American Countries. *Am. J. Polit. Sci.* 29(3):539
- Cingranelli DL, Richards DL. 1999a. Measuring the Pattern, Level, and Sequence of Government Respect for Human Rights. *Int. Stud. Q.* 43(2):407–17

- Cingranelli DL, Richards DL. 1999b. Respect for Human Rights after the End of the Cold War. *J. Peace Res.* 36:511–34
- Clark A. 2001. *Diplomacy of Conscience: Amnesty International and Changing Human Rights Norms*. Princeton: Princeton University Press
- Clark AM, Sikkink K. 2013. Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures. *Hum. Rights Q.* 35(3):539–68
- Claude RP. 1976. Empirical Comparative Rights Research: Some Preliminary Tests of Development Hypothesis. In *Comparative Human Rights*, ed. RP Claude, pp. 51–76. Baltimore: Johns Hopkins University Press
- Conrad CR. 2012. Divergent Incentive for Dictators: Domestic Institutions and (International Promises Not to) Torture. *J. Confl. Resolut.* 56(5):1–34
- Conrad CR, Ritter EH. 2013. Treaties, Tenure, and Torture: The Conflicting Domestic Effects of International Law. *J. Polit.* 75(2):397–410
- Coppedge M, Gerring J, Lindberg SI, Teorell J, Pemstein D, et al. 2014. V-Dem: A New Way to Measure Democracy. *J. Democr.* 25(3):159–69
- Crabtree CD, Fariss CJ. 2015. Uncovering Patterns Among Latent Variables: Human Rights and De Facto Judicial Independence. *Res. Polit.* 2(3):1–9
- Creamer CD, Simmons BA. 2015. Ratification, Reporting, and Rights: Quality of Participation in the Convention against Torture. *Hum. Rights Q.* 37(3):579–608
- Dancy G, Fariss CJ. 2017. Rescuing Human Rights Law from Legalism and its Critics. *Hum. Rights Q.* forthcoming:

- Dancy G, Michel V. 2016. Human Rights Enforcement From Below: Private Actors and Prosecutorial Momentum in Latin America and Europe. *Int. Stud. Q.* 60(1):173–88
- Dancy G, Montal F. 2017. Unintended Positive Complementarity: Why International Criminal Court Investigations Increase Domestic Human Rights Prosecutions. *Am. J. Int. Law.* forthcoming:
- Dancy G, Sikkink K. 2012. Treaty Ratification and Human Rights Prosecutions: Toward a Transnational Theory. *NYU J. Law Int. Polit.* 44(3):751–90
- Davenport C. 1995. Multidimensional Threat Perception and State Repression: An Inquiry into Why States Apply Negative Sanctions. *Am. J. Polit. Sci.* 39:683–713
- Davenport C. 1996. “Constitutional Promises” and Repressive Reality: A Cross-National Time-Series Investigation of Why Political and Civil Liberties Are Suppressed. *J. Polit.* 58:627–54
- Davenport C. 2007. *State Repression and the Domestic Democratic Peace*. Cambridge: Cambridge University Press
- Davis KE, Kingsbury B, Merry SE. 2015. The Local-Global Life of Indicators: Law, Power, and Resistance. In *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*, eds. KE Davis, B Kingsbury, SE Merry, pp. 1–24. Cambridge: Cambridge University Press
- De Neufville JI. 1986. Human rights reporting as a policy tool: An examination of the State Department Country Reports. *Hum. Rights Q.* 8(4):681–699
- DeMars WE. 2005. *NGOs and Transnational Networks: Wild Cards in World Politics*. London: Pluto Press

- Dezalay Y, Garth BG. 2002. Legitimizing the New Legal Orthodoxy. In *Global Prescriptions: The Production, Exportation, and Importation of a New Legal Orthodoxy*, eds. Y Dezalay, BG Garth. Ann Arbor: University of Michigan Press
- Diamond L. 2011. Why Democracies Survive. *J. Democr.* 22(1):17–30
- Donnelly J, Howard RE. 1988. Assessing national human rights performance: A theoretical framework. *Hum. Rights Q.* 10(2):214–248
- Elliott MA. 2011. The institutional expansion of human rights, 1863-2003: A comprehensive dataset of international instruments. *J. Peace Res.* 48(4):537–46
- Fariss CJ. The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance. *Br. J. Polit. Sci.*
- Fariss CJ. 2014. Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability. *Am. Polit. Sci. Rev.* 108(2):297–318
- Fein H. 1995. More Murder in the Middle: Life-Integrity Violations and Democracy in the World, 1987. *Hum. Rights Q.* 17(1):170–91
- Finnemore M, Toope SJ. 2001. Alternatives to “Legalization”: Richer Views of Law and Politics. *Int. Organ.* 55(3):743–58
- Foucault M. 1980. Truth and Power. In *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*, ed. C Gordon. New York: Pantheon
- Fukuyama F. 2006. *The End of History and the Last Man*. Simon and Schuster
- Gastil R. 1980. *Freedom in the World: Political Rights and Civil Liberties, 1980*. New Brunswick, NJ: Transaction Books
- Giannone D. 2010. Political and Ideological Aspects in Measurement of Democracy: the Freedom House Case. *Democratization.* 17(1):68–97

Goldstein JS. 2012. *Winning the War on War: The Decline of Armed Conflict Worldwide*.

New York: Dutton

Goldstein RJ. 1986. The limitations of using quantitative data in studying human rights abuses. *Hum. Rights Q.* 8(4):607–627

Grant S. 2011. Amnesty's Achievements: Some Reflections on the Early Years. In *Amnesty International 50 Years - Reflections and Perspectives*, eds. W de Jonge, B McGonigle Leyh, A Mihr, van Troost, pp. 11–21. Utrecht, Netherlands: SIM

Green M. 2001. What we talk about when we talk about indicators: current approaches to human rights measurement. *Hum. Rights Q.* 23(4):1062–1097

Greenhill B. 2010. The Company You Keep: International Socialization and the Diffusion of Human Rights Norms. *Int. Stud. Q.* 54(1):127–45

Gurr TR, Scarritt JR. 1989. Minorities rights at risk: A global survey. *Hum. Rights Q.* 11(3):375–405

Hafner-Burton EM. 2014. A Social Science of Human Rights. *J. Peace Res.* 51(2):273–86

Hafner-Burton EM, Ron J. 2009. Seeing Double: Human Rights Impact through Qualitative and Quantitative Eyes. *World Polit.* 61(2):360–401

Hafner-Burton EM, Tsutsui K. 2005. Human Rights in a Globalizing World. The Paradox of Empty Promises. *Am. J. Sociol.* 110(5):1373–1411

Hafner-Burton EM, Tsutsui K. 2007. Justice Lost! The Failure of International Human Rights Law to Matter Where Needed Most. *J. Peace Res.* 44(4):407–25

Harff B, Gurr TR. 1988. Toward Empirical Theory of Genocides and Politicides: Identification and Measurement of Cases Since 1945. *Int. Stud. Q.* 32(3):359

- Hathaway OA. 2002. Do Human Rights Treaties Make a Difference? *Yale Law J.* 11(8):1935–2042
- Henderson C. 1982. Militaray Regimes and Rights in Developing Countries: A Comparative Perspective. *Hum. Rights Q.* 4:110
- Hibbs Jr. DA. 1973. *Mass Political Violence: A Cross-National Causal Analysis.* New York: Wiley
- Hill DW Jr, Inglett J. 2015. A Unified Measure of Women’s Rights
- Hollyer JR, Rosendorff BP. 2012. Do Human Rights Agreements Prolong the Tenure of Autocratic Ratifiers? *NYU J. Int. Law Polit.* 44(3):791–811
- Hopgood S. 2013. *The Endtimes of Human Rights.* Ithaca, NY: Cornell University Press
- Howard R. 1983. The Full-Belly Thesis: Should Economic Rights Take Priority Over Civil and Political Rights? Evidence from Sub-Saharan Africa. *Hum. Rights Q.* 5(4):467–490
- Howse R, Teitel R. 2010. Beyond Compliance: Rethinking Why International Law Really Matters. *Glob. Policy.* 1(2):127–36
- Humana C. 1983. *World Human Rights Guide.* London: Hutchinson
- Kalantry S, Getgen JE, Koh SA. 2010. Enhancing Enforcement of Economic, Social, and Cultural Rights Using Indicators: A Focus on the Right to Education in the ICESCR. *Hum. Rights Q.* 32:
- Keck ME, Sikkink K. 1998. *Activists beyond Borders : Advocacy Networks in International Politics.* Ithaca, N.Y.: Cornell University Press
- Keith LC. 2012. *Political Repression.* Philadelphia, PA: University of Pennsylvania Press

- Keys B. 2010. Congress, Kissinger, and the Origins of Human Rights Diplomacy. *Dipl. Hist.* 34(5):823–851
- Landman T. 2004. Measuring human rights: principle, practice and policy. *Hum. Rights Q.* 26(4):906–931
- Landman T, Carvalho E. 2009. *Measuring Human Rights*. New York: Routledge
- Linzer DA, Staton JK. 2015. A Global Measure of Judicial Independence, 1948–2012. *J. Law Courts.* 3(2):223–56
- Lupu Y. 2013a. Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements. *Int. Organ.* 67(3):469
- Lupu Y. 2013b. The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects. *Am. J. Polit. Sci.* 57(4):912–25
- Lupu Y. 2015. Legislative Veto Players and the Effects of International Human Rights Agreements. *Am. J. Polit. Sci.* forthcoming:
- Mason TD, Krane DA. 1989. The Political Economy of Death Squads: Toward a Theory of the Impact of State-Sanctioned Terror. *Int. Stud. Q.* 33(2):175–98
- McKinlay RD, Cohan AS. 1976. Performance and Instability in Military and Nonmilitary Regime Systems. *Am. Polit. Sci. Rev.* 70(3):850–864
- McNitt AD. 1988. Some Thoughts on the Systematic Measurement of the Abuse of Human Rights. In *Human Rights: Theory and Measurement*, ed. DL Cingranelli, p. 89. St. Martin's Press
- Mitchell NJ, McCormick JM. 1988. Economic and Political Explanations of Human Rights Violations. *World Polit.* 40(4):476–498

- Munck GL, Verkuilen J. 2002. Conceptualizing and Measuring Democracy: Evaluating Alternative Indices. *Comp. Polit. Stud.* 35(1):5–34
- Murdie A, Bhasin T. 2011. Aiding and Abetting: Human Rights INGOs and Domestic Protest. *J. Confl. Resolut.* 55(2):163–91
- Neumayer E. 2005. Do International Human Rights Treaties Improve Respect for Human Rights? *J. Confl. Resolut.* 49(6):925–53
- Neyer J, Wolf D. 2005. The analysis of compliance with international rules: definitions, variables and methodology. In *Law and Governance in Postnational Europe*, eds. M Zürn, C Joerges, pp. 40–64. New York: Cambridge University Press
- Norden E. 1966. American Atrocities in Vietnam. *Liberation*
- Park H. 1987. Correlates of Human Rights: Global Tendencies. *Hum. Rights Q.* 9:405–13
- Pinker S. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Penguin Books
- Poe SC. 1990. Human Rights and US Foreign Aid: A Review of Quantitative Studies and Suggestions for Future Research. *Hum. Rights Q.* 12(4):499
- Poe SC, Tate NC. 1994. Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis. *Am. Polit. Sci. Rev.* 88:853–900
- Poe SC, Tate NC, Keith LC. 1999. Repression of the Human Right to Personal Integrity Revisited: A Global Crossnational Study Covering the Years 1976-1993. *Int. Stud. Q.* 43(2):291–315
- Powell EJ, Staton JK. 2009. Domestic judicial institutions and human rights treaty violation. *Int. Stud. Q.* 53(1):149–74

- Raustalia K, Slaughter A-M. 2002. International Law, International Relations and Compliance. In *Handbook of International Relations*, ed. W Carlnaes, pp. 538–58. Sage Publishing
- Regan PM, Henderson EA. 2002. Democracy, threats and political repression in developing countries: Are democracies internally less violent? *Third World Q.* 23(1):119–36
- Rejali D. 2007. *Torture and Democracy*. Princeton: Princeton University Press
- Rodley SN. 2011. Amnesty International’s Work on Physical Integrity--A Personal Reflection. In *Amnesty International 50 Years - Reflections and Perspectives*, eds. W de Jonge, B McGonigle Leyh, A Mihr, van Troost, pp. 51–74. Utrecht, Netherlands: SIM
- Ron J. 2000. Savage Restraint: Israel, Palestine, and the Dialectics of Legal Repression. *Soc. Probl.* 47(4):445–72
- Rosga A, Satterthwaite ML. 2009. The Trust in Indicators: Measuring Human Rights. *Berkley J. Int. Law.* 27(2):253–315
- Rubin BR, Newberg PR. 1980. Statistical analysis for implementing human rights policy. In *The Politics of Human Rights*. New York: NYU Press
- Rummel RJ. 1976. *The Dimensionality of Nations Project*. Ann Arbor, MI: ICPSR Codebooks
- Satterthwaite ML. 2016. Coding Personal Integrity Rights: Assessing Standards-Based Measures against Human Rights Law and Practice. *N. Y. Univ. J. Int. Law Polit.* 48(forthcoming):

- Schoultz L. 1981. U. S. Foreign Policy and Human Rights Violations in Latin America: A Comparative Analysis of Foreign Aid Distributions. *Comp. Polit.* 13(2):149
- Schrodt PA. 2014. Seven deadly sins of contemporary quantitative political analysis. *J. Peace Res.* 51:287
- Scoble H, Weisberg L. 1981. Problems of Comparative Research in Human Rights. In *Global Human Rights: Public Policies, Comparative Measures and NGO Strategies*, eds. V Nanda, JR Scarritt, G Shepherd. Boulder, CO: Westview Press
- Scott JC. 1998. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press
- Sikkink K. 2011. *The Justice Cascade: How Human Rights Prosecutions Are Changing World Politics*. New York: W.W. Norton & Company
- Simmons B. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press
- Simmons BA, Danner A. 2010. Credible Commitments and the International Criminal Court. *Int. Organ.* 64(2):225–56
- Spirer HF. 1990. Violations of Human Rights-How Many?: The Statistical Problems of Measuring Such Infractions Are Tough, but Statistical Science Is Equal to It. *Am. J. Econ. Sociol.* 49(2):199–210
- Stohl M, Carleton D, Johnson SE. 1984. Human Rights and U.S. Foreign Assistance from Nixon to Carter. *J. Peace Res.* 21(3):215–25
- Stohl M, Carleton D, Lopez G, Samuels S. 1986. State violation of human rights: Issues and problems of measurement. *Hum. Rights Q.* 8:592

- Stohl M, Lopez GA, eds. 1986. *Government Violence and Repression: An Agenda for Research*. New York: Greenwood Press
- Taylor C, Jodice D. 1983. *World Handbook of Political and Social Indicators, 3rd Ed.* New Haven, CT: Yale University Press
- The Anti-Apartheid Movement. 1966. Annual Report. London
- UN General Assembly. 1984. *The Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment*
- Vreeland JR. 2008. Political Institutions and Human Rights: Why Dictatorships enter into the United Nations Convention Against Torture. *Int. Organ.* 62(1):65–101
- Wood R, Gibney M. 2010. The Political Terror Scale (PTS): A Re-introduction and a Comparison to CIRI. *Hum. Rights Q.* 32(2):367–400
- Zunino M. 2011. Releasing Transitional Justice from the Technical Asylum: Judicial Reform in Guatemala seen through Techne and Phronesis. *Int. J. Transitional Justice.* 5:99–118

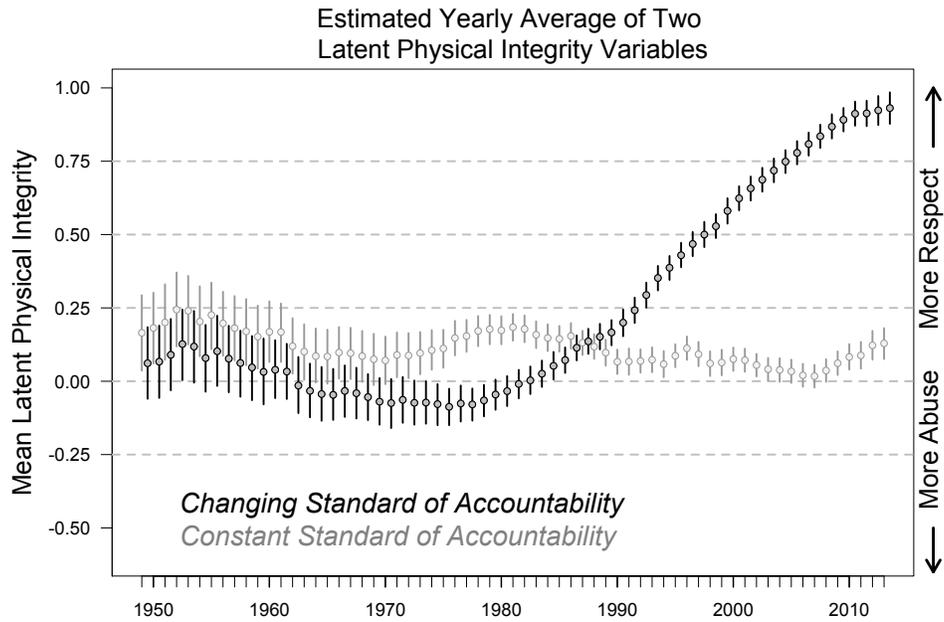


Figure 1: The panel displays yearly mean and credible intervals for latent physical integrity estimates from the changing standard latent variable model and the constant standard latent variable model (Fariss, 2014). Only the latent variable estimates that assume a changing standard of accountability show improvement for either type of country-year.

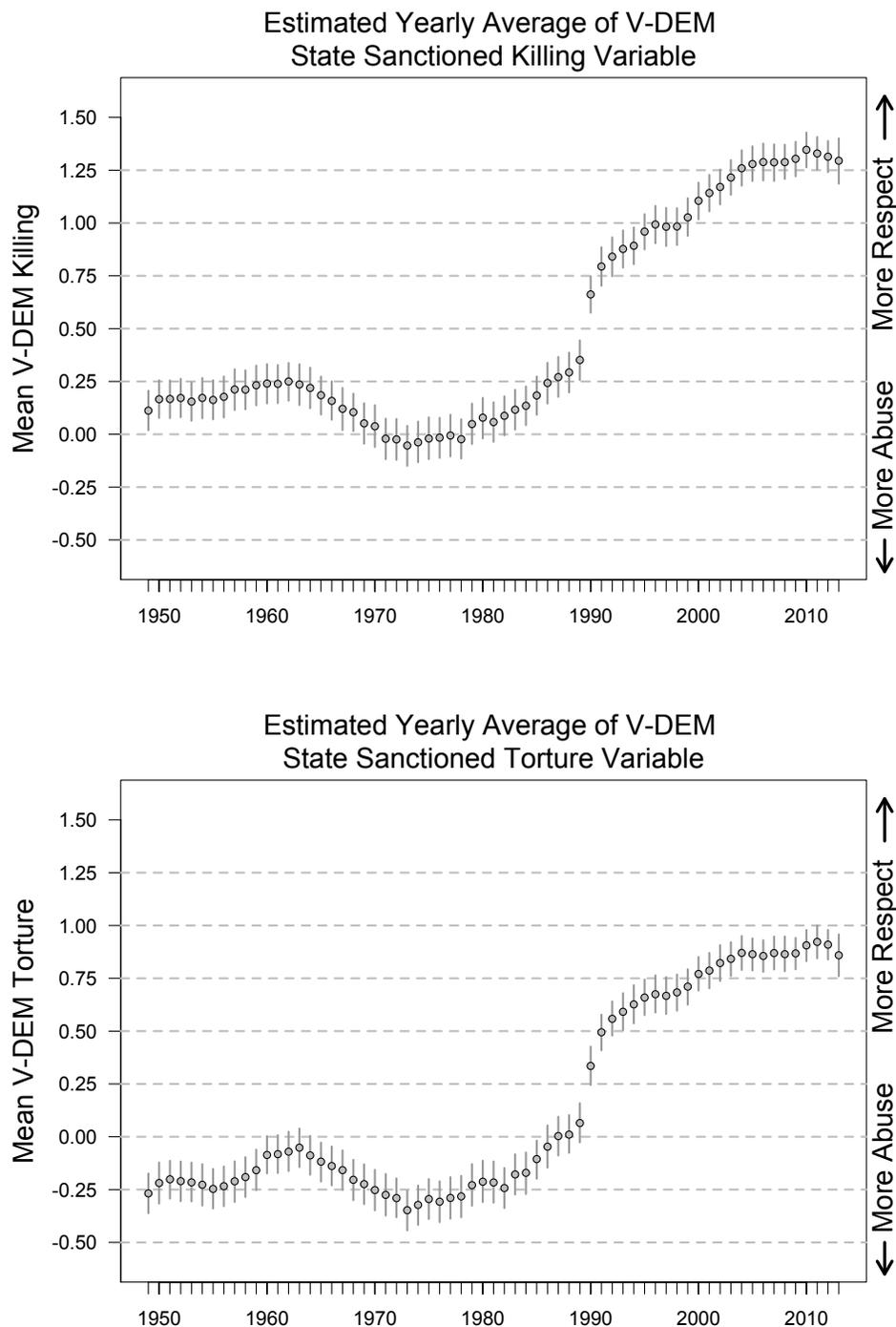


Figure 2: The yearly average for the two expert-coded V-DEM physical integrity variables from 1949-2013 (Coppedge et al., 2014; Pemstein, Tzelgov and Wang, 2015), which is the same time period available for the most recent update of latent human rights variable. What should be clear from this visualization, is a very similar upward trend in human rights respect after the end of Cold War. These positive trends coincide with the pattern of increasing respect for human rights found with the new data from Fariss (2014).

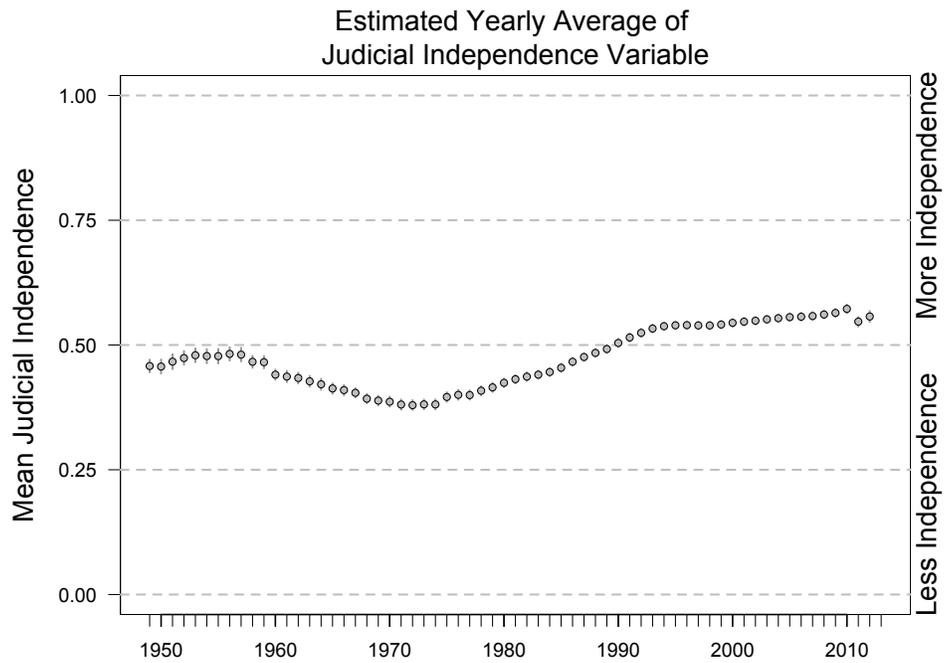


Figure 3: The yearly average for the Linzer and Staton Judicial Independence data 1949-2012 Linzer and Staton (2015). This positive trend coincides with the pattern of increasing respect for human rights found with the new data from Fariss (2014).

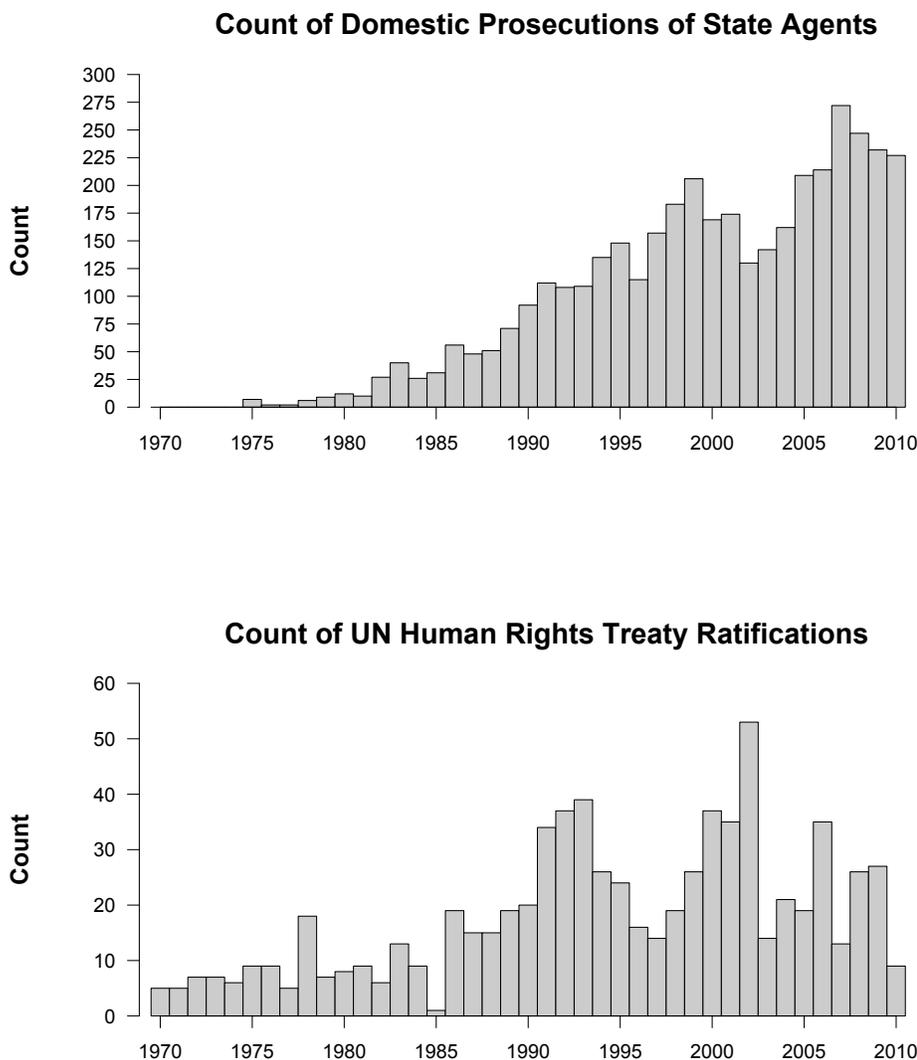


Figure 4: The top panel displays yearly count of the number of domestic criminal prosecutions for human rights abuses (1970-2010). The lower panel displays the yearly number of country ratifications of United Nations Human Rights Treaties (1970-2010). These positive trends coincide with the new view of increasing respect for human rights found with the new data from Fariss (2014). Data are taken from Dancy, Geoff, Francesca Lessa, Bridget Marchesi, Leigh A. Payne, Gabriel Pereira, and Kathryn Sikkink. 2014. “The Transitional Justice Research Collaborative: Bridging the Qualitative-Quantitative Divide With New Data.” Available at www.transitionaljusticedata.com.

- Coppedge, Michael, John Gerring, Stefan I. Lindberg, Jan Teorell, Daniel Pemstein, Eitan Tzelgov, Yi ting Wang, Adam Glynn, David Altman, Michael Bernhard, M. Steven Fish, Allen Hicken, Kelly McMann, Pamela Paxton, Megan Reif, Svend-Erik Skaaning and Jeffrey Staton. 2014. "V-Dem: A New Way to Measure Democracy." *Journal of Democracy* 25(3):159–169.
- Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability in Human Rights Documents." *American Political Science Review* 108(2):297–318.
- Pemstein, Daniel, Eitan Tzelgov and Yi ting Wang. 2015. "Evaluating and Improving Item Response Theory Models for Cross-National Expert Sur." *Varieties of Democracy Institute Working Paper 1* pp. 1–53.