

# Human Rights Reports and the Organizations that Produce Them

Carly Potz-Nielsen, Robert Ralston, Thomas R. Vargas

University of Minnesota

This draft: April 11, 2017

## **Abstract**

Human rights reports produced by Amnesty International and the United States' State Department are used by researchers and policymakers alike to understand the human rights conditions within countries across time. A recent debate in the human rights literature has suggested that the coding of such reports carries substantial potential for bias. We test the argument that popular human rights monitoring reports are subject to both data-generating and standards-based “information effects,” and argue that these “information effects” can be understood as a measurement problem that biases inferences made from human rights reports. We use two computer-assisted text analysis methods—event data coding and Latent Semantic Analysis—to analyze the manifest features of the reports that give clues as to how documents change (or do not) over time. We also interview individuals familiar with the drafting and editing process at both Amnesty International and the State Department. Our findings suggest that organizational effects, namely “procedural biases” make it difficult to use these reports in the manner in which they are often used. This has important implications for scholars interested broadly in looking “under the hood” of popular sources for human rights and political violence-based data.

# 1 Introduction

Amnesty International and the US State Department produce annual human rights reports which are used by researchers and policymakers to understand human rights abuses within countries across time. State Department country-year reports, for example, are used by judges to adjudicate asylum requests in the U.S., as principal documents in the legal systems of Canada, Australia, and New Zealand, and by foreign ministries, press, and business interests as a source of information. Their expansion beyond their original purpose to help Congress determine aspects of American aid relationships illustrates the growing reliance of communities on these documents. Similarly, it is not clear that Amnesty International Reports were meant for anything other than their own advocacy purposes and yet these documents are also used in court decisions, particularly in Europe and cited in other reports detailing human rights abuses, like the State Department.

In an effort to quantify the information provided by the country-year reports, scholars have used human coders to generate indices representing the occurrence of human rights violations. The Political Terror Scale (PTS) and the Cingranelli-Richards Human Rights Data Project (CIRI) have been deployed in quantitative models attempting to explain levels of human rights abuses over time. For instance, previous work understanding political regime survival and repression ([Escribà-Folch, 2013](#)), impacts of international nongovernmental organizations ([Murdie and Bhasin, 2011](#); [Murdie and Peksen, 2014](#)), international commitments and compliance ([Kelley, 2007](#); [Kim and Sikkink, 2010](#)), domestic development of economic and political institutions ([Wood, 2008](#); [Powell and Staton, 2009](#)), and trends in human rights violations ([Clark and Sikkink, 2013](#)) all utilize these quantitative measures.

However, researchers have speculated that the content and structure of these reports—for example, their length and focus, as well as the quality of the information therein—often changes across countries and time ([Clark and Sikkink, 2013](#); [Fariss, 2014](#)). Less attention has been drawn to the fact that these documents are products of distinct organizations that research, write, and publish their reports using different standards, norms, and processes

over time. Under such conditions, it becomes difficult to distinguish between real changes in the level of human rights abuses world-wide, shifting meaning with respect to what counts as a ‘human right,’ as well as the preferences and politics of the organizations that produce the reports. This raises concerns about what information is being relayed by indices derived from these reports, particularly with respect to the *true* levels of human rights, within countries and worldwide, over time.

In this paper, we use computer-assisted text analysis methods to uncover evidence of a organization-level process determining the information relayed by the reports. We examine the country-year reports produced by Amnesty International and the US State Department from 1980 - 2012 using these text analytic techniques and incorporate insights from interviews with individuals at Amnesty International and the State Department who are familiar with the reporting, editing, and publishing process of these reports. Our contribution builds upon efforts to “look under the hood” of these popular reports to understand the information generating process within them over time, as well as the features of the text that may bias inferences that are drawn from the documents (Bagozzi and Berliner, 2016). In particular, we find evidence of three processes active in the data: (1) temporal variation in the amount of human rights violations reported by each organization; (2) temporal variation in how the information is gathered and interpreted; and (3) temporal and inter-organizational variation in the topical focus of each report.

The remainder of this paper is divided into six sections. The first section lays out the background of the country-year reports created by Amnesty International and the State Department. The second section outlines prior literature that attempted to identify and measure bias in the reports, focusing specifically on the impact of changing attention and capacity of organizations and global shifts in the understanding of human rights. The third and fourth sections present our two methods, event coding via the TABARI software (Schrodt, 2014b) and Latent Semantic Analysis, and discuss their output as they relate to changes in the amount and content of information over time, respectively. In the fifth section, we

leverage our conversations with current and former State Department and Amnesty officials to highlight the impacts of procedure on the information contained in the reports. Finally, we conclude by suggesting that using the country-year reports as indicators of *over time* change in the occurrence of human rights abuse is inappropriate given the process by which the organizations produce them.

## 2 Background: Human Rights Reports

Amnesty International and the State Department have produced country-year reports since the late 1970s. Both organizations use these reports to provide a snapshot of pertinent human rights abuses for a given country and year, with an eye on relaying crucial information to their respective audiences. While each follow the same basic editing process, key differences in their procedures help to highlight the different information contained in the reports. Table 1 summarizes such differences, which are the focus of this section.

Amnesty uses their human rights reports to inform their members about the worst human rights abuses in different countries in a given year. Amnesty places a priority on primary sources of information, but also uses local media reports and secondary sources as evidence in their reports. As a nongovernmental organization funded by fees and donations of members worldwide, the scope of countries covered in their annual reports has been limited by both its agenda and its capacity. Put another way, these reports, while a massive undertaking, are the product of an NGO for a particular audience: their members. Reports vary in their attention and detail to countries over and across time due to pressures from their audience (Ron et al., 2005), as well as logistical constraints faced when compiling the reports. For instance, Amnesty reports must fit into a booklet that is no more than 25,000 words long.

Country-year reports produced by the US State Department rely on the expertise of foreign service officers stationed in the relevant country, information from domestic and international NGOs, specifically Amnesty International or Human Rights Watch, and interna-

Table 1: **Procedural Differences in Organizations**

	<b>Amnesty International</b>	<b>U.S. State Department</b>
<i>Who determines the focus of the report?</i>	Secretary General and Regional Offices	Congress/Senior Officials in Washington
<i>How are violations organized?</i>	If there are country-specific violations that do not fit into issue area, then deviate (rarely occurs)	Violations must be fit within the structure defined yearly by instruction cable from D.C.
<i>How are the reports drafted?</i>	Initial draft → Thematic Experts and Lawyers → Editorial Staff → Fact Verification → Regional Research Manager → Copy Editorial Team → Translation → Year End Country Developments Incorporated → Publication	Embassies → Editors in D.C. → Country Desk Officers and DRL Regional Editors → Senior Staff* → Embassies
<i>What sources do the reports usually use?</i>	Personal network and secondary sources within country	Embassy: personal network and NGOs Editors: Internet and D.C.-based NGOs
<i>Is there a word count?</i>	Word count (500 - 2500)	No, but each country report must address every section.
<i>Where does the organization receive its funding?</i>	Donations	Congressional budget
<i>When are the reports drafted and released?</i>	Sept - January; Released late February	July - February; Usually released within first half of the year.

\* Depending on whether country report is of interest or controversial. For instance, reports on China and Russia are viewed by senior management every year.

tional news sources. These reports are mandated by Congress<sup>1</sup> and were originally intended to inform on governments receiving aid from the United States. Despite their popularity, Congress remains the principal audience for State Department reports. Further, because the State Department is an executive agency, it is sensitive to the political considerations of government, meaning that political and editorial attention varies over countries and time for

<sup>1</sup>Created in the Foreign Assistance Act of 1961 and the Trade Act of 1974.

strategic reasons. For example, our interviews suggest that certain “high interest” countries (like Russia, China, and the Middle East) get more attention than others at higher levels of the State Department.

In a dated, but insightful, description of the State Department reports, de Neufville (1986, p. 689-690) describes the individuals who produce these reports as neither ‘political appointees nor ordinary bureaucrats.’ de Neufville goes on to note that “they are somewhat independent of the particular politics of an administration, not only because of their career status, but also because they have professional expertise and values. The Country Reports team is increasingly willing and able to resist pressures to slant the Reports for political purposes.” (ibid.). de Neufville’s description points to two possible organizational dynamics that impact how information is generated and processed in the State Department reports: political dynamics, like a strategic focus on a particular country, and the impact of organizational personnel as knowledge producers who vary over time. Previous analyses of the State Department reports in the late-1970s through the late-1990s have found, compared to Amnesty International, the State Department’s reports have been marked, at times, by harsher coverage of leftist regimes and better coverage of U.S. trading partners and allies (Poe et al., 2001).<sup>2</sup>

Amnesty International and State Department reports **are not** written with an eye on future coding by human rights scholars.<sup>3</sup> Moreover, scholars using reports for anything other than they are intended should understand the differences by which Amnesty International and the State Department produce these reports. This includes paying attention to the reports’ diverging aim, audience, and the particular incentives and priorities Amnesty and State face when reporting on particular countries (Bagozzi and Berliner, 2016). We argue that any scholar using these documents should move beyond the assumption that these reports are simply relaying “objective” or “ground truth” information.<sup>4</sup>

---

<sup>2</sup>Our interviews with individuals familiar with the process through which the State Department produces reports confirm these dynamics are at play.

<sup>3</sup>This was echoed across all of our interviews.

<sup>4</sup>See, for example, the Political Terror Scale’s words of caution with respect to this aspect of the reports:

### 3 The Information Generation Process

Though qualitative evidence suggests increasing levels of protection of human rights worldwide, this trend does not bear out in measures quantifying human right abuses. Clark and Sikking (2013) suggest that this is because in using quantitative measures, like the Political Terror Scale (PTS) and the Cingranelli-Richards Human Rights Data Set (CIRI), researchers fail to separate out two data generating processes:<sup>5</sup> (1) the process by which human rights violations occur and (2) the process by which the scores are created.<sup>6</sup> Clark and Sikking argue that uneven reporting over time due to increases in the capability to collect information and attention to abuses biases quantitative measures (2013, p. 546). Clark and Sikking define these issues of data quality in human rights reporting as “information effects,” which are “patterns in the data stemming from the process of information collection and interpretation, rather than the process that actually gives rise to human rights violations or their mitigation” (2013, p. 540). Changes in the quality and amount of information over time implies that violations unobservable in previous years become observable as access increases and cost of gathering information decreases. This is understood as the driving force masking the trend suggested by qualitative evidence. In effect, yearly reports are not comparable over time because they are measuring human rights violations based off of systematically differing levels of information.<sup>7</sup>

Richards (2016) critiques Clark and Sikking’s use of document length, measured by word count, as a proxy for the level information contained within the reports. Richards points

---

<http://www.politicalterroryscale.org/archive/SDvsAI/>

<sup>5</sup>This is a concern that is not unique to human rights data, and that has long been central to the literature on political violence, similarly concerned with how violent incidents or casualties, are reported (Danzger, 1975; Caliendo et al., 1999; Davenport and Ball, 2002; Davenport, 2010; Weidmann, 2014, 2016; Vargas et al., 2016).

<sup>6</sup>The PTS provides a measure of threats to physical integrity rights, used to measure annual state-based political violence against a country’s population. CIRI scores creates four separate indicators of physical integrity abuses that takes into account the frequency at which violations occur. Both indices are coded from the country-year reports produced by the US State Department and Amnesty International.

<sup>7</sup>This would be akin to measuring public opinion towards China since the 1970s, but gradually expanding the sample from large metropolitan areas in the early years to including all other areas in the latter years, while treating each sample as comparable.

to the use of human coders, in the case of the construction of CIRI specifically, to argue that human coders can substantively analyze documents; increases in word count would not deceive a human coder with regard to the 'level' of abuse (2016, p. 490). In addition, CIRI's operational procedure includes recoding previous years as standards of human rights evolve addressing changing norms on the face of the coding (ibid.). However, updating coders' understandings provide no value added if the information for those violations do not exist in the underlying source documents. Wood and Gibney (2010) echo this point by claiming that there is not enough information in the source documents—that is, the Amnesty International and State Department country-year reports—for the level of detail that CIRI claims to represent.

Fariss (2014) derives a measurement model to account for “changing standards of accountability” that he argues produce the differing content of reports over time. If Amnesty and the State Department have changed the scope and depth of their investigation and reporting of human rights abuses over time, and have expanded their definition of what constitutes a “human rights abuse,” then these reports will contain different levels of specificity, accuracy, and completeness over time. His model takes into account the changing incentives of organizations to gather more accurate information, broaden the kinds of information reported on, and pressure governments through traditional human rights advocacy tactics like naming and shaming (ibid., 312). Fariss argues that these processes bias conclusions drawn from the quantitative scores developed based on organization reports, and his model attempts to separate out the underlying trend in human rights abuses from that of the reporting process of the CIRI score over time. Changing standards of accountability mean that trends within human rights documents, such as the Amnesty International's and the State Department's annual reports, have been masked by the changing context in which the reports were written in the first place, resulting in scholars missing decreasing levels repression (ibid.).

We address concerns related to changing standards of accountability and move beyond Fariss' analysis by subjecting the reports themselves to examination, rather than focusing on

the processes that transform reports to indices. Changes in the process by which these reports are compiled suggests that while the macro-trends (Fariss, 2014; Bagozzi and Berliner, 2016) introduce bias when comparing human rights over time, there may be an additional layer of procedural bias varying over time at the organizational level. In other words, the processes through which the reports are produced, from the editorial process and decision-making structure of the organizations to the directives handed down by the Amnesty and State Department hierarchies shapes the content that appears within the reports. This, we argue, introduces a level of *procedural bias* thus far unexplored by the literature. Our analysis seeks to parse the process by which information is collected and consolidated from the process by which events occur. To be clear, we are not seeking to produce a new measure of human rights violations, but instead to provide a new framework for examining how the information generation process impacts the reports over time.

## 4 Methods: Amount and Type of Information

We examine information within the Amnesty International and US State Department country-year reports along two dimensions: how much information is produced, and what type of information is provided. We look for evidence of organizational differences over time using country reports on Human Rights Practices for Argentina, Guatemala, and Mexico, made available by Fariss et al (2015). First, we use the event-coding software TABARI (Schrodt, 2014b), which leverages the structure of the text to capture instances of abuse. Rather than employing a 'bag of word' approach, TABARI uses shallow parsing to identify 'who did what to whom' from the syntactic structure of relevant sentences and to code these actions into relational events.<sup>8</sup> Second, to address the possibility of broader structural changes within the reports over time we employ Latent Semantic Analysis, which treats the reports as a "bag of words" to identify latent meaning, to look at the change in the type of information

---

<sup>8</sup>More information regarding TABARI and the process by which we coded events can be found in Appendix 8.1

contained in the documents.

We are interested in differences between organizations as well as differences between countries. In our interviews (discussed below), individuals with experience at various stages of the publication process at both Amnesty and the State Department noted that there is an emphasis for different country reports to be structurally similar so that they are comparable in a given year. Since our cases are all within the same region, we also can assume that they are reviewed by the same regional editors for each year in both the State Department and Amnesty. This has the added benefit of holding constant the potential for regional effects, following Hafner-Burton and Ron's (2013) suggestion for more "regionally attuned" human rights scholarship.

The country-year reports are produced to relay information on human rights abuses and, problematically, the indices coded from them specifically prioritize the frequency of a particular abuse.<sup>9</sup> We conceptualize the "amount of information" within a document as the number of events coded within a given document. We define an "event" as an instance of abuse detailed in the reports which specifies an actor, an action, and a target of abuse. As the length of each report changes over time due to changing format or demands of publication, we expect the amount of events to vary accordingly. Amnesty International consistently produces shorter reports than the State Department.<sup>10</sup> Therefore, in order to compare the amount of information both over time and across organizations, we weight the amount of events by the word count. This avoids the misinterpretation that the State Department reports are consistently "more informative" than the Amnesty reports if the focus is solely on word counts. While the State Department produces longer reports, the reports often contain more "filler" information than instances of abuse.

By looking instead at the density of information (number of events / number of words), we are able to compare the amount of information in the reports per word. This means

---

<sup>9</sup>Specifically, the CIRI Physical Integrity Score is coded based on how often a particular right is violated. More information in Appendix 8.2.

<sup>10</sup>Amnesty has a word count of 500-2500 words, no such minimum or maximum exists for State Department. Graphs presenting word count can be found in Appendix 8.2.

that increases in the length of reports due to the addition of descriptive information rather than instances of abuse will be accounted for in our analysis. If there are information effects in the documents, we should see a consistent trend in the level of information relayed over time. This would indicate a systematic shift in the ability and interest of the reporting organizations. Variation from year-to-year could indicate an absolute change in instances of abuse “on the ground” within a country or, alternatively, procedural changes such as changes in the desired use of particular examples in the reports or in the number of specific abuses deemed necessary to characterize the human rights situation in that country.<sup>11</sup> We leverage our cross-country comparison and background information regarding the reports to attempt to differentiate between the two.

Latent Semantic Analysis (LSA) evaluates the similarity in the latent meaning of the country-year reports, which allows us to examine whether there have been changes in the type of information relayed by the documents. Our interviews with former State Department officers confirm that the State Department relies on other organizations to produce the reports. Both through formal meetings, where Amnesty may lobby the State Department to focus on a type of abuse, to informal personal relationships to confirm information and instances of abuse, these organizations are intertwined beyond their focus on the same object of study. We should then expect there to be a high degree of similarity between the organizational reports. To gauge the degree of this similarity, we assess the correlation between the reports produced by both organizations and make inferences about changes in the documents’ ‘true meaning’ and their relationship to each other.

If there are changing standards of accountability over time, then reports should become more standardized in their meaning with less variation in their similarity. That is, if reporting agencies are converging on common meanings of standards, yearly change in the type of information that gets reported should become more stable over time. However, if it is the

---

<sup>11</sup>An interviewee from Amnesty International, for example, noted that particular types of violations do not make it into the report in a given year, particularly if the violations are well outside the boundaries of the general theme that year.

case they are not systematically changing, then there should be little variation in meaning over time. LSA can capture this relationship by tapping into the semantic structure of the documents and identifying the distance between any two documents in the semantic space.

## 5 Findings

By looking at the amount and type of information across reports and organizations, we are able to contribute to the debate between Clark and Sikkink and Richards and offer a new process for understanding changes in the production of these reports over time. First, we argue that while Clark and Sikkink are right to point to changes in availability of information over time, their use of word count may overstate the extent to which it is biasing our understanding of human rights abuse. We find initial support for Richards' claim that CIRI coders were able to take these information effects into account. However, we also find evidence of changing standards over time, supporting Fariss, which cannot be accounted for by human coders. Our results indicate strong differences between organizations suggesting a refinement to Fariss' explanation: organizational dynamics.

### 5.1 The Impact of Information Effects

The concern for Clark and Sikkink is that changes in the amount of information within the reports translates into higher reported instances of abuse over time, irrespective of the "ground-truth." Clark and Sikkink use word count as a proxy to look at the change in the amount of information and find evidence of systematic change. Comparing the word count of the country-year reports to the number of events identified by TABARI, we see similarity in the trends over time. Figure 1 presents the over time variation in the amount of events for Guatemala, Argentina, and Mexico (1980 - 2012). In this formulation, more events can be understood as more information.

The first graphs highlight the difference across organizations of the trend over time and

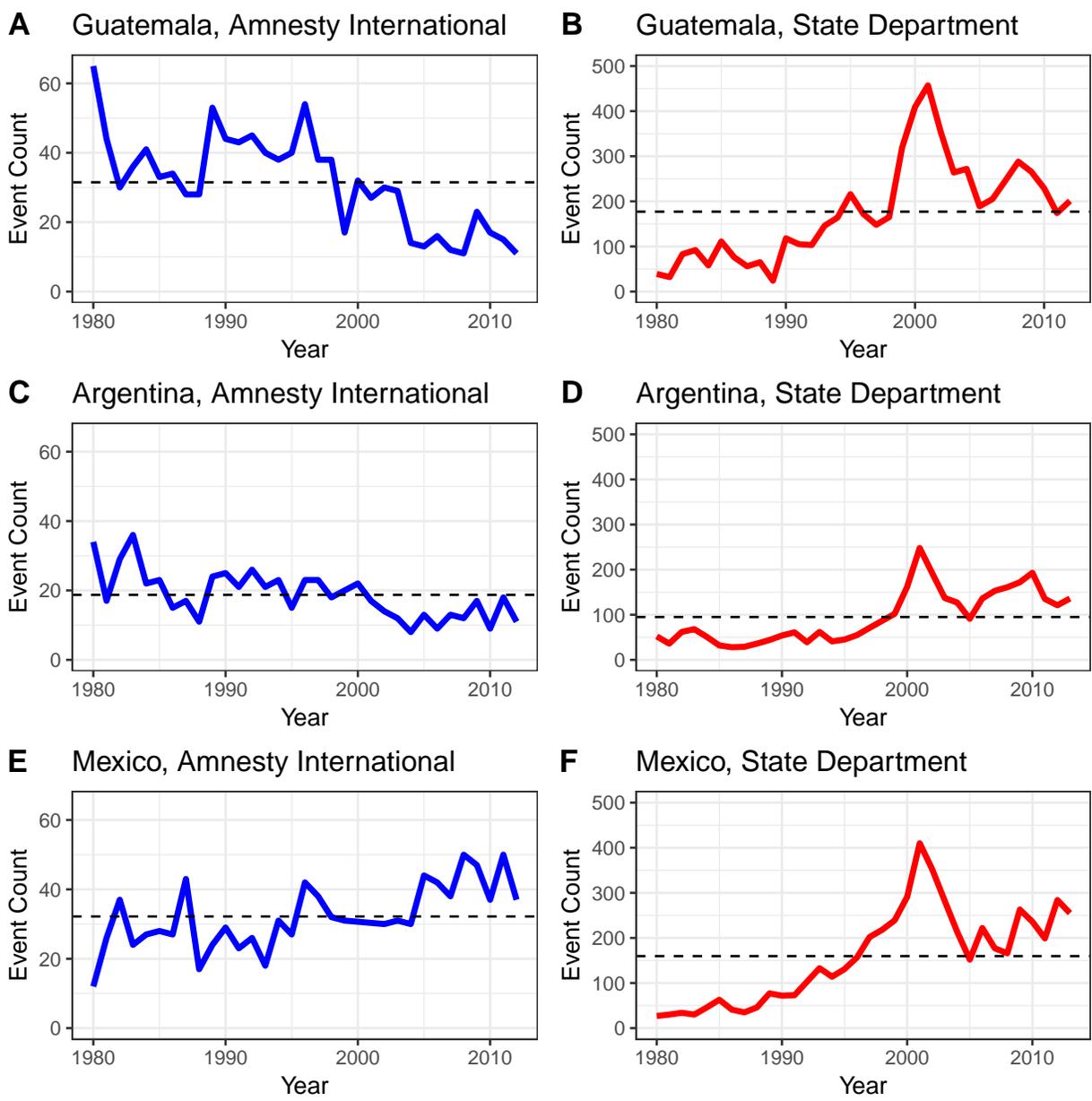


Figure 1: Raw Event Counts Over Time

the amount of events. Amnesty reports across our three countries show no clear trend; any increase in events in a given year is offset by subsequent falls. By contrast, in the late-1990s the State Department reports start to show similar trends. Notably, there is a sharp increase in events followed by a large decrease in the early 2000s and subsequent larger fluctuations in the total number of events from year to year. The isolation of this trend to the State Department reports suggests an organizational-specific driving factor; we find it

highly unlikely that the number of “real” abuses on the ground jumped/peaked in all three countries around the same time without Amnesty International capturing a similar trend.

The difference in the absolute amount of events further highlights the difference between organizations. Amnesty’s maximum number of events is just above 60 in Guatemala at the beginning of the 1980s but trends much lower than that for most of the time series. The mean number of events over this time period (indicated by the dashed line) for the State Department are, on average, 114 events higher than their Amnesty counterparts.<sup>12</sup> However, this trend is not a given moving forward. For instance, in 2010 a decision was made to cut down the length and breadth of the reports as they were becoming too unwieldy for embassies to complete.<sup>13</sup>

To compare the amount of information across organizations over time, we divide the event count for each document by the word count, presented in Figure 2. Longer documents have more space to convey information than shorter ones. By weighting the number of events by word count, we can get a sense of the proportion of events compared to less specific information detailing broader country conditions.

We see a general reversal from the graphs in Figure 1: Amnesty exhibits a higher number of events per word than the State Department.<sup>14</sup> This reversal suggests that document length is a driving force behind raw event count and supports Richards’ claim: information effects may not be problematic with respect to the CIRI score coding because humans are able to see beyond sheer increases in document length and sift through the ‘fluff’ to code the number of abuses reported. Therefore, changes in word count appear to have no significant bearing on the coding of abuses.

The crux of Clark and Sikkink and Richards’ debate centers around two key claims: (1) that the State Department reports contain more information over time; and (2) that more

---

<sup>12</sup>Descriptive statistics are provided in Appendix 8.2

<sup>13</sup>Based upon an interview with the Editor-in-Chief of the State Department’s reports

<sup>14</sup>On average, the State Department reports less events per word than Amnesty for all three countries. The difference between the means are all statistically significant at the .05 level. For every mutual 300 word increase, Amnesty has, on average, approximately 1 more event than the State Department.



Figure 2: Event Counts / Word Counts

words per report is indicative of more information. Clark and Sikkink argue that more words in reports leads to more information (Richards 2016, p. 487). While Clark and Sikkink are correct that as the number of words in the document increases, so too does the amount of

information—measured as the number of events mentioned in the text—it is not clear that the magnitude of the increased information would mask the underlying trend of human rights abuses over time. This supports Richards’ rebuttal: information effects may exist, but they do not result in biased scores.<sup>15</sup>

More broadly, we again see clear differences across organizations supported by differences in their aims and process. Amnesty International focuses on particular individual level abuses, in line with their organizational focus and intended audience of INGOs and transnational civil actors. Meanwhile, the State Department, writing primarily for Congress and policymakers more generally, includes information regarding the state of the country, focusing on broad patterns of abuses and conditions of violations.

## 5.2 Changes in Content over Time

Fariss outlines an alternative source of bias within the State Department and Amnesty International reports, highlighting the role of larger shifts in the focus the reports and organizations over time. While human coders may be able to distill abuses from changes in the amount of information, changes in content produces a different hurdle; coders cannot go back and code for an abuse that was not recorded at the time. To examine the change in content we leverage both the changes in the verbs defining an event, sorting verbs into pre-determined categories for interpretation, and use LSA to inductively identify change in the latent meaning of the reports.

We extract the operative verb of an event identified by TABARI and categorize it as one of four verb categories: physical integrity rights, civil/political rights, economic/social rights, and other verbs related to human rights.<sup>16</sup> Figure 3 provides the count of events falling in each of the four verb categories.<sup>17</sup>

First, Amnesty International privileges physical integrity rights in their reporting. While

---

<sup>15</sup>As additional evidence, we provide in Appendix 8.2 the density of information compared with the CIRI Physical Integrity Index and find no clear relationship in their movement over time.

<sup>16</sup>This is, for all intents and purposes, a different way of modelling pre-determined topics over time by distilling Bagozzi and Berliner’s (2016) topics into crude categories for analytic convenience. Table of verbs

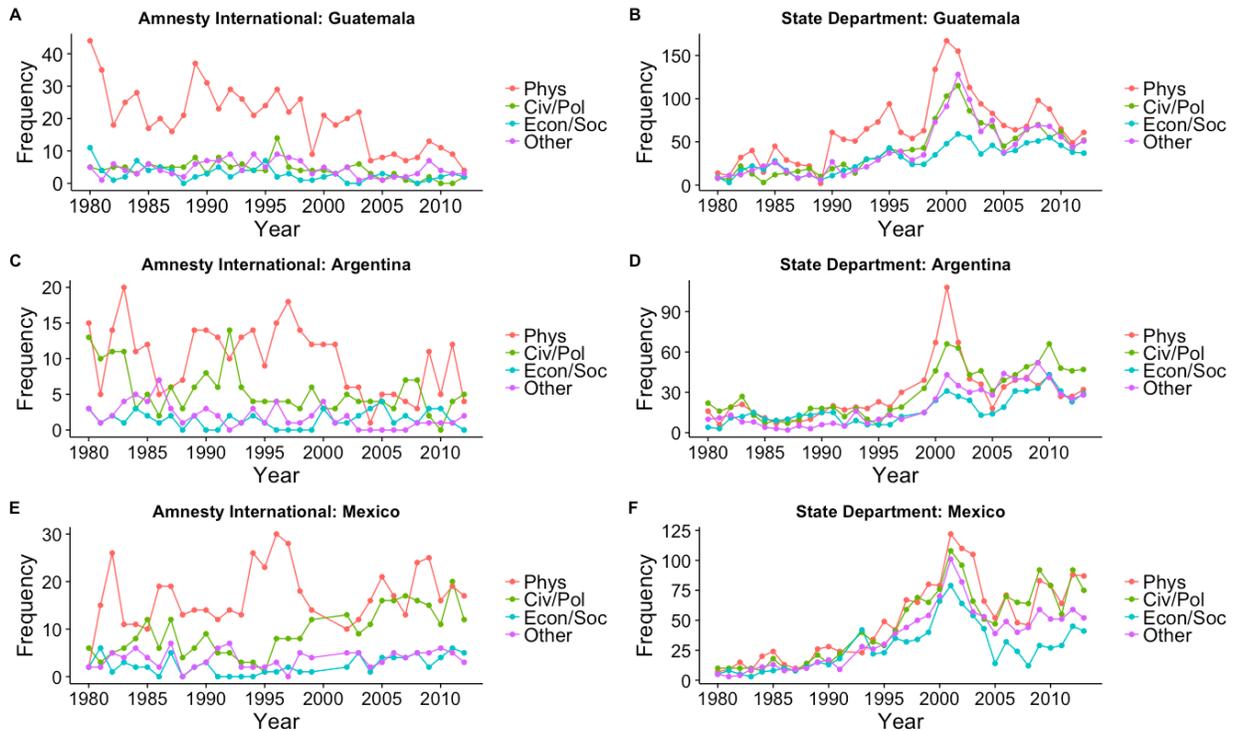


Figure 3: Verb Category Use Over Time

this attention decreases overtime, their priority remains on reporting instances of physical abuse aligning with their origin as an advocacy network against torture. By comparison, the State Department not only spreads its attention more evenly across the categories, but also exhibits a consistency over time and across countries. This aligns with the structure and process of State Department reports. The organization places a high degree of importance on the comparability of reports across countries mandating a specific structured format via the “Instruction Cable” sent to all embassies at the beginning of the compiling process each year. Differences across organizations suggest an additional process to Fariss’ changing standards of accountability; beyond global shifts in focus and definition of human rights abuses there are organizational-specific trends in content.

Figure 3 tells us little about the content of the reports themselves as it forces events into

---

provided in Appendix 8.1

<sup>17</sup>Graphs presenting proportion of events in each verb category appear in Appendix 8.2. It should be noted that the verb categories are chosen specifically for ease of illustration. We are interested in their change over time; the categories are a simple and consistent way to illustrate how focus changes, not the absolute focus of the reports.

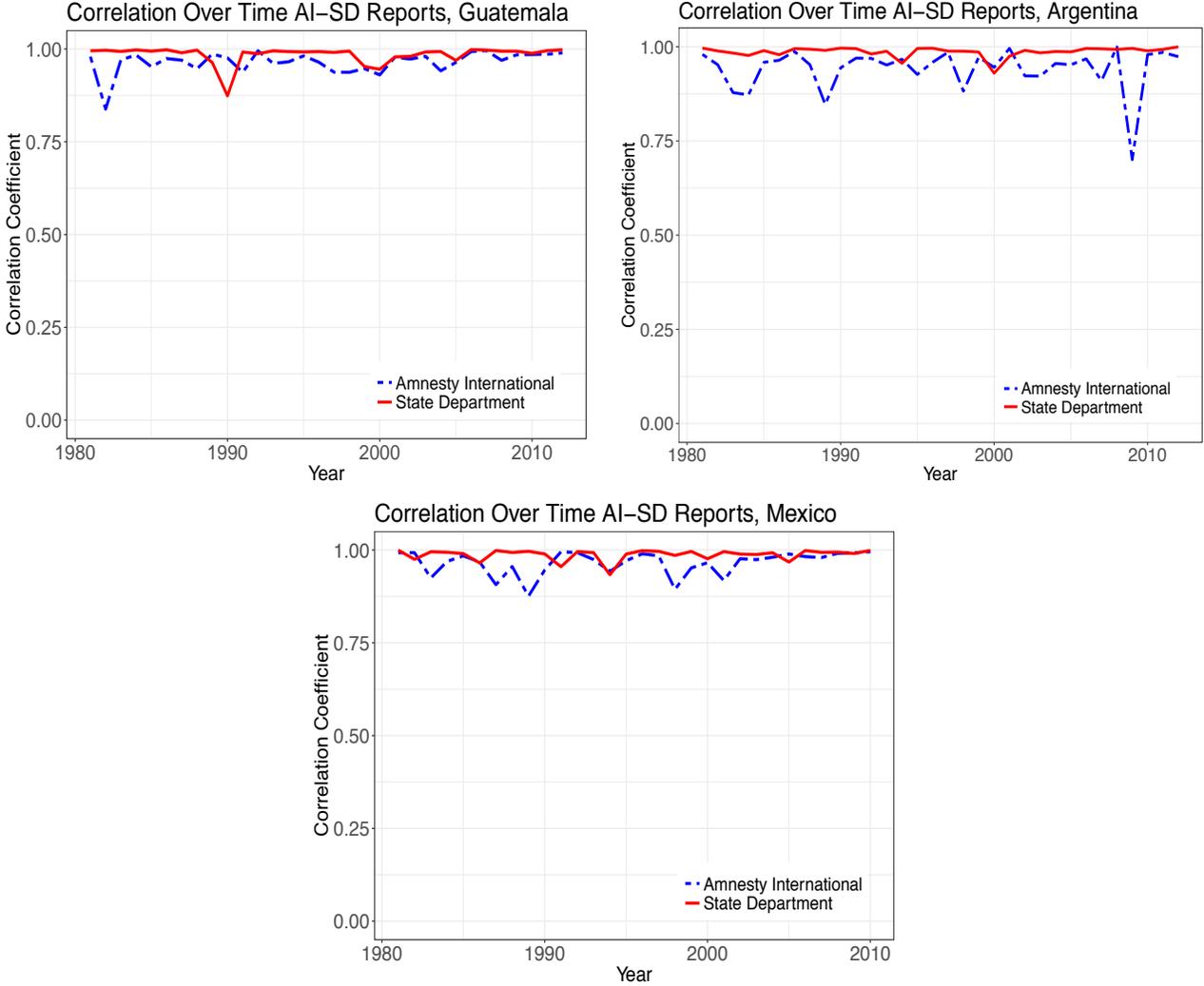


Figure 4: Correlation within Organizational Reports Over Time

four broad categories. To inductively identify the content of the reports and its change over time, we use LSA to map the change in content over time. If the documents were identical from year-to-year, in terms of their latent content, they would occupy the same position in the semantic space and the correlation between them would be 1. If the documents were entirely distinct, they would occupy diametrically opposed positions in the latent semantic space, and the correlation between them would approach zero. By mapping the similarity of documents in the latent space, understood as the correlation of a report at time  $t-1$  with a report at time  $t$ , we can examine whether there are systematic changes in the type of information reported by each organization. Figure 4 displays the the correlation between

reports across time for each country, within the same organization.

Organizational changes in the content of reports, leads us to expect documents to be closely related with a consistent trend over time. Specifically, with respect to Fariss' argument, we would expect that variation from year-to-year *and* from organization-to-organization would decrease as the correlation between and across reports increases over time. The difference in the variability of latent similarity across organizations supports our earlier conclusion: State Department reports are more similar over time as they are produced by a more institutionalized and stable process. By contrast, Amnesty shows more variation in the type of information relayed within its country-year reports.

We can get a sense of what is contained within the latent semantic space, i.e. what topics are being focused on from year-to-year, by examining the top 10 loading words for Amnesty and State Department reports for 1980, 1990, 2000, and 2010 (Table provided in Appendix 8.3). This provides us with rough snapshots of the position of the documents within the latent semantic space. As a general trend, we note that the State Department reports concentrate more heavily on institutions, such as courts, the military, and law, whereas Amnesty International reports focus most heavily on people and particular violations, such as killings, torture, and disappearances.

In support of Fariss' argument, we observe shifts in the focus of the documents, particularly for Amnesty reports. The most recent Amnesty International reports on Argentina, place words such as indigenous, communities, and provinces at the heart of the reports, suggesting a renewed interest in typically marginalized groups. A similar trend can be picked up in Guatemala where the 2010 report includes 'women' among the top 10 loading words. This is not to suggest, however, that the AI reports from 1980 and 2010 are diametrically opposed; it is likely that the reports are quite similar, but that the focus has changed slightly to reflect changes in human rights concerns over time. Taken along with the patterns of verbs across the organizations, this provides additional evidence for systematic shifts in the content of reports. Though not quantifiable, there is also additional evidence indicating that

Amnesty reports are more fluid and responsive to changes in focus and content, highlighting the driving impact of organizational-differences.

## 6 Discussion: Organizational Effects

Our analysis of these reports and our interviews with individuals involved in creating the reports point to a common theme: these reports, produced by both Amnesty and the State Department, should be used carefully, especially when viewed over time. Interviewees from both Amnesty International and the State Department expressed unease at the notion that these reports should be used to track human rights abuses over time. Rather, as both our text analyses and our interviews suggest, these documents are the products of the organizations that produced them.

While others have highlighted macro-level changes in human rights scores derived from these reports over time, we point to potential meso-level trends that occur prior to the publishing of the documents. Several organizational facets of these reports are worth mentioning. First, the reports are never meant to represent “ground truth” in each country, nor are they exhaustive accounts. This is especially important when thinking about the *lack* of events in a particular context—this is not necessarily indicative of zero human rights abuses of that kind that year, but rather it could be a function of the organizational decisions that privilege some information over other information. Moreover, interviewees suggested that if there is no change in a particular situation over time, then the abuses often get dropped from subsequent reports, despite the abuse continuing. In the case of Amnesty, an instance was relayed to us in which Amnesty wanted to draw attention to police brutality in a particular country. However, since they did not see themselves making any headway with regard to changing the practices, they dropped that issue from their reports. Additionally, because of the sensitive nature of some of the abuses, like cases of torture, it may be the case the primary sources want to report to Amnesty, but not make their claims public in fear of

retribution. In both examples, Amnesty would intentionally not mention either abuse due to organizational constraints, even though they have knowledge of its occurrence.

On the other hand, it may be the case that an abuse that was occurring in previous years was not included in the reports until relevant to the issue areas. The State Department is especially sensitive to changes in focus as they are beholden to the structure of the reports, which is predetermined in Washington. A clear example is the focus on abuses targeted at women in the 1993 country reports(Shattuck, 1993). The instruction cable for that year directed embassies to pay special attention to instances of rape by government officials to be included in the section on torture, issues regarding voting rights to be covered in the section on the right to choose one's own government, and the specific request to focus on the trafficking of women within the reports. From these reports, it would be incorrect to conclude that these abuses against women were not occurring in previous years.

While most scholars would agree that these reports are not representative of the actual number of abuses on the ground, there is still a reliance on them, and the scores coded from them, to illustrate change over time. Underlying this understanding is the assumption of consistency in the way that these reports are produced. Both Clark and Sikkink and Fariss highlight the dangers of this assumption, arguing that the reports are pressured by systematic changes. We focus on a different level, that of the organization, and argue that changes in procedure cause variation in the comparability of the reports over time. An example of this is the restructuring of the State Department's reports in 2010. In order to reduce the time and space dedicated to these reports, embassies were instructed to only report one pertinent event per type of abuse, instead of documenting different instances<sup>18</sup>. The amount and type of information relayed in the reports directly changed due to adjustments in the procedural instructions.

---

<sup>18</sup>Personal Interview, Interviewee 3

## 7 Conclusion

Social scientists often depend upon reports produced by governmental and nongovernmental organizations in order to inform their research. For example, regulatory bodies produce reports on compliance with various international rules and norms and governments produce white papers that detail their foreign policy priorities and concerns. These reports are often used to inform theories regarding how states behave, how leaders think, and the political dynamics of different organizations. By analyzing the amount, density, and type of information found within documents as well as the relationship between documents, researchers can begin to unpack the ways in which other important series of documents are produced. We offer a process by which to think about information *over time*, how it is produced and how organizational processes impact our understanding of change.

With respect to our analysis of human rights reports for Argentina, Guatemala, and Mexico, we found evidence that information changes over time as a function of changes in the process by which these reports are generated. Our analysis indicates an absolute increase in the amount of information relayed over time, but this does not necessarily translate to the addition of new, substantive information regarding instances of abuse. However, we also find that, in examining the *type* of information relayed over time, we find support for changes in the focus of the reports, differing across organizations highlighting the impact of organization-specific processes.

With respect to the debate in the pages of *Human Rights Quarterly*, we conclude that Clark and Sikkink are right to point to the possible impact of information effects, but that they may be overstating the magnitude. In highlighting the ability of human coders to see beyond sheer increases in amount, we support Richards' argument that information effects are not systematically biasing CIRI score downward. However, changes in content impact our understanding of human rights violations over time and are driven not just by global shifts in focus, as suggested by Fariss, but by organizational procedures and characteristics.

More broadly, our interviews suggest the danger of conceptualizing country-year reports

as accurate representations of “ground-truth.” Both Amnesty and State Department decide upon yearly issues of focus that then determine what is included and emphasized within the country-reports for that year. The need for internal consistency and comparability limits what events and types of abuses are covered and in what way. While we are not the first to suggest that using country-reports as ground-truth is misleading, we seek to address not how accurate the information is, but how consistent it is over time. In shifting the focus towards changes in the information relayed by the reports, we provide an indication of the variation in the bias over time resulting from the reporting process of organizations. Future research should further theorize and empirically test for organizational dynamics within these reports. For example, there may be ways in which organizational processes and dynamics can be introduced and controlled for in models that seek to examine changing human rights standards over time.

## 8 Appendix

### 8.1 TABARI Dictionary Development

TABARI (Schrodt, 2014b) is a program that uses pattern recognition to automatically code events from text. An event is defined as “*who did what to whom.*” TABARI reads individual sentences and identifies event patterns using actors, verbs, and phrases that are contained within the accompanying dictionaries. This produces a data set wherein each observation is an event, defined as an actor-verb-target set.

Automated techniques generally have the advantage of stability, reproducibility, and accuracy (Weber, 1990; Schrodt, 2014a). Stability in the coding is often cited as the main motivator for using automated over hand-coding techniques. As Schrodt (2014a, 5) notes: “The stability of machine coding is 100% because the machine will always code the same text in the same manner. This is particularly useful when a time series is being maintained for a number of years.” Second, our coding process is reproducible. This means that any scholar interested in replicating our results, given the same data, dictionary, and steps, will be able to reproduce our measures. Additionally, this same process can be extended to other applications focusing on other countries, or other contexts. Moreover, unlike human coders, TABARI conforms exactly, and instantaneously, to coding changes and rules. The machine, in other words, will do precisely what it is told, in a consistent and reproducible manner. Finally, while recent efforts on “human computation” techniques—that give simple tasks to thousands of trained but non-expert coders—are promising, they remain expensive because you have to train and pay coders per task.

The benefit of using TABARI also extends to its event coding scheme. TABARI provides not only raw event counts, but uses dictionaries to code and then place events into different predetermined categories. We leverage this aspect of TABARI’s coding scheme by asking TABARI to code for three different types of violations: Civil and Political Rights, Economic, Social, and Cultural Rights, and Physical Integrity Rights.<sup>19</sup> The two main types of recog-

---

<sup>19</sup>We also create a category called “other,” which catches frequently-occurring verbs in the reports that

nized human rights are Civil and Political Rights (originating from the first portion of the UDHR in 1948) and Economic, Social, and Cultural Rights (found in the UDHR, but also the International Covenant on Economic, Social and Cultural Rights (ICESCR)). In other words, TABARI provides us with details not only about the number of events found within each report—a measure of concrete information—but also the type of event.

TABARI codes events using source-actor-target patterns with three separate dictionaries: actors, agents, and verbs. TABARI then uses these dictionaries to parse sentences and relies on the grammatical structure to locate events within the sentence (illustrated in Figure 5). Dictionary development is a key factor to correctly coding events from the text. TABARI relies on sparse parsing, as opposed to full syntactical analysis, meaning that it focuses on locating the proper nouns, verbs, and direct objects of the verb phrase. TABARI does not by itself learn from the text, it merely reads the structure as defined by location of the words specified by the different dictionaries. We focus on the most frequently occurring verbs and actors within the text, both for the sake of time efficiency and also to reduce any possible bias introduced by hand-coding texts from different documents.

To create the actor dictionary, we first used Python’s Natural Language Toolkit (NLTK) to extract a list of proper nouns from the text of the documents for each country. Then, we ran this list of proper nouns through an online text analysis program (<https://www.online-utility.org/text/analyzer.jsp>) that counts the frequency of each word and selected only those proper nouns that occurred ten or more times within each country’s set of documents. These were then taken as the most frequently occurring actors and used to develop the actor dictionary. To generate our agent dictionary, we use the ICEWS agent dataset, directly importing their agent list. Finally, we created our verb dictionary using the results of the topic models from Berliner and Bagozzi (2016). We took all the resulting FREX (frequently occurring and exclusive) verbs from topic models run on the entire corpus of documents, cross-nationally, and sorted them into convenient verb categories for descriptive

---

do not fit in one of the three categories in a straightforward manner.

Date: 31 Dec 80

Record : AIARG###

In December 1980 prisoners transferred in handcuffs from Unidad 6, Rawson, to Unidad 9, La Plata were severely beaten by military personnel.

Coded events:

801231 PRV991 PRV625 103 \*null\* MILITAR --- PRI ---

Select: N)ext P)arsing M)odify R)ecode A)utocode O)ther Q)uit ->

```
<clause>
0 IN <prep> IN_
1 DECEMBER
2 1980 <number> 1980
3 PRISONERS <actor> PRI
4 TRANSFERRED
5 IN <prep> IN_
6 HANDCUFFS
7 FROM
8 UNIDAD <litrl> UN
9 6 <number> 6
10 , <comma> ,
11 RAWSON
12 , <comma> ,
13 TO
14 UNIDAD <litrl> UN
15 9 <number> 9
16 , <comma> ,
17 LA
18 PLATA
19 WERE <auxil> WERE_
20 SEVERELY
21 BEATEN <verb> BEAT
22 BY <by_> BY_
23 MILITARY <actor> MILITAR
24 PERSONNEL
</clause>
```

Figure 5: Example of TABARI Sentence Parsing

purposes, shown below in Table 2. It is important to note that these categories are only used for description and are not used for analysis regarding the types of events coded from the texts. Instead, their purpose is to give a illustrative sense of what verbs are contained within the event data and how it maps over time.

Table 2: **TABARI Verb Categories**

Violation Type	Verbs
Physical Integrity Rights	abduct, assassin, beat, disappear, execut, harass, imprison, injur, interrog, kidnap, kill, murder, prostit, rape, sexual, threat, tortur, arm, attack, attacked, attend, bomb, fight, forc
Civil/Political Rights	access, appoint, bail, bond, control, corrupt, demonstr, guard, investig, occupi, reform, deport, detain, detaine, prohibit, protect
Economic/Social Rights	commiss, confisc, degrad, develop, disabl, discrimin, displac, divorc, employ, exploit, guarante, marri, protest, provid, regist, strike, trade
Other	amend, await, board, practic, print, secur, see, strip, studi, traffick, train, act, assist, depart, recommend, refus

### TABARI Event Examples

#### Physical Integrity Rights:

In December 1980 *prisoners* transferred in handcuffs from Unidad 6, Rawson, to Unidad 9, La Plata were severely **beaten** by *military personnel*.

#### Civil/Political Rights:

On 4 April the chief of police in Buenos Aires province stated that *Angel Romano* had been officially **detained**.

#### Economic/Social Rights:

They had gone on hunger-**strike** on 10 September when *Congress* failed to find time to consider a draft law which would have enabled them to be conditionally released while their cases were being reviewed.

#### Other:

After his detention was acknowledged *Miguel Andel De Ple* was still **refused** visits from his lawyer.

#### “Fluff”:

The search for children reported as missing following the abduction of their parents made some progress during the year.

## 8.2 TABARI Output

Within this section are additional descriptive statistics (Table 3) and graphs providing a visual comparison of the raw event counted extracted by TABARI to the word count provided by Fariss (2015) meta-data (Figure 6 - 11). Additionally, we map the change in the

the density of events (Event Count/ Word Count) to the change in the countries Physical Integrity Score provided by CIRI.

Table 3: **Descriptive Statistics**

Country Report	Word Count				Event Count			
	Mean	St. Dev	Min	Max	Mean	St. Dev	Min	Max
Guatemala SD	11700	6786	2804	32320	177.1	109.3	24	457
Guatemala AI	1778	619	787	3607	31.48	13.74	11	65
Argentina SD	7001	3219	3092	15690	94.91	59.08	28	248
Argentina AI	1226	444.9	489	2783	18.73	6.861	8	36
Mexico SD	11890	6966	3626	30340	158.7	104.67	27	410
Mexico AI	2040	385	1070	2924	32.19	9.481	12	50

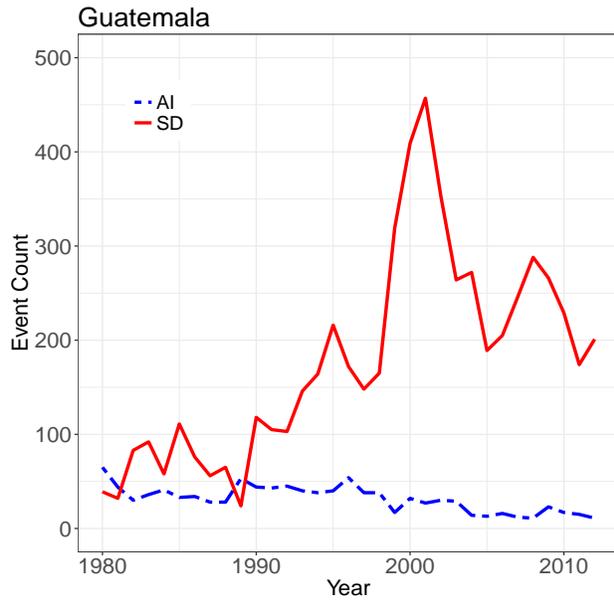


Figure 6: GTM Event Counts

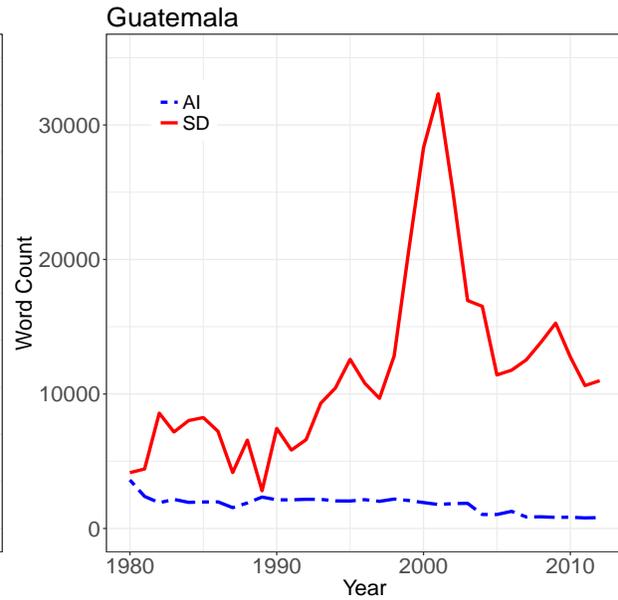


Figure 7: GTM Word Count

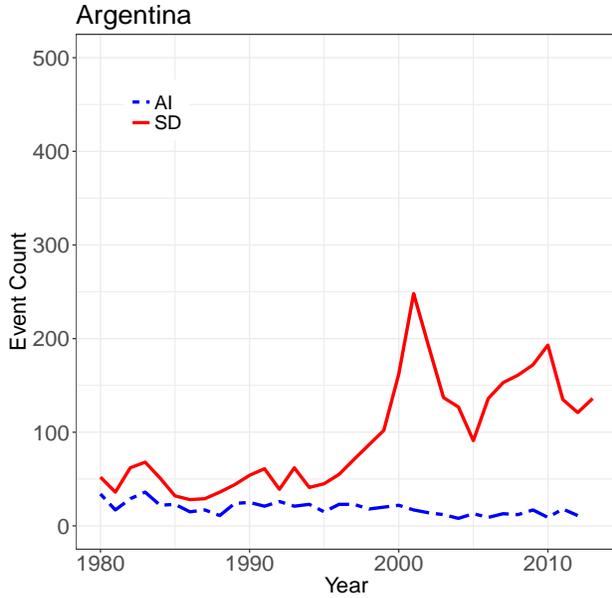


Figure 8: ARG Event Counts

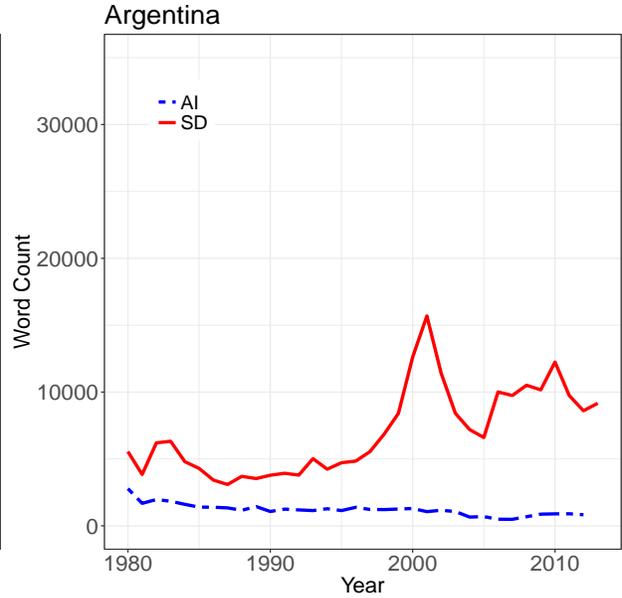


Figure 9: ARG Word Count

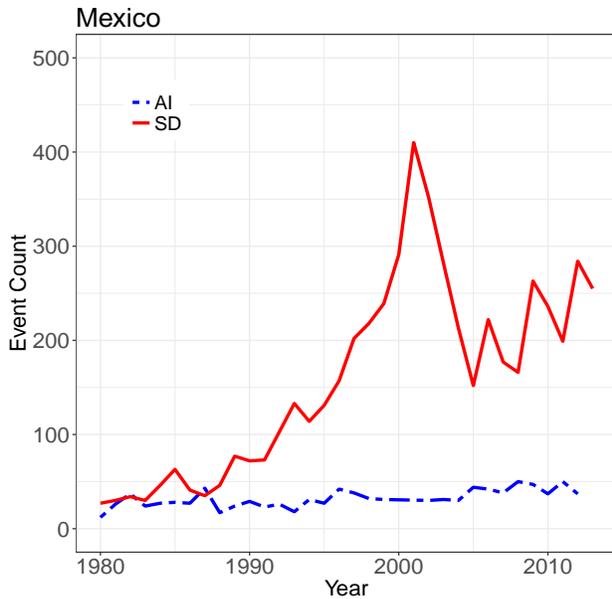


Figure 10: MEX Event Counts

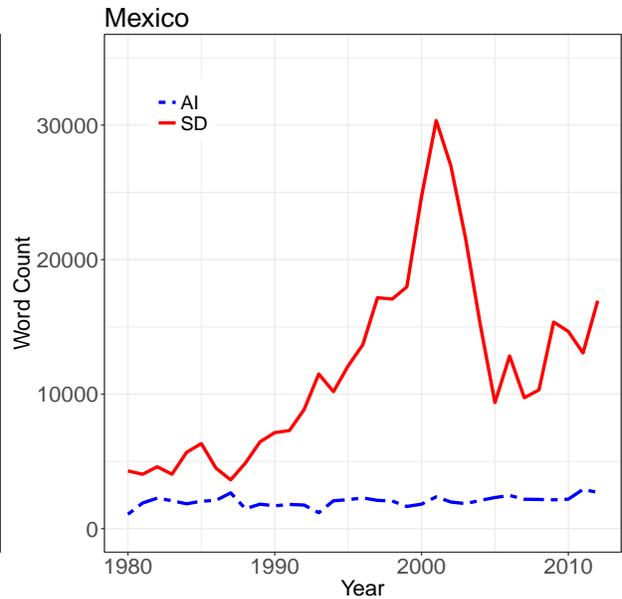


Figure 11: MEX Word Count

Figure 12 provides an over time comparison of the CIRI Physical Integrity Score and amount of events, per 100 words, for each year. The CIRI Physical Integrity Index is an additive index constructed from the Torture, Political Imprisonment, Disappearance, and Extrajudicial Killing indicators. These indicators take on the value of 0, if the abuse occurred frequently within a year, 1 if the abuse occurred occasionally, and 2 if the abuse

did not occur. The Physical Integrity Index ranges from from 0, indicating no government respect for these four rights, to 8, full government respect for these four rights.<sup>20</sup>

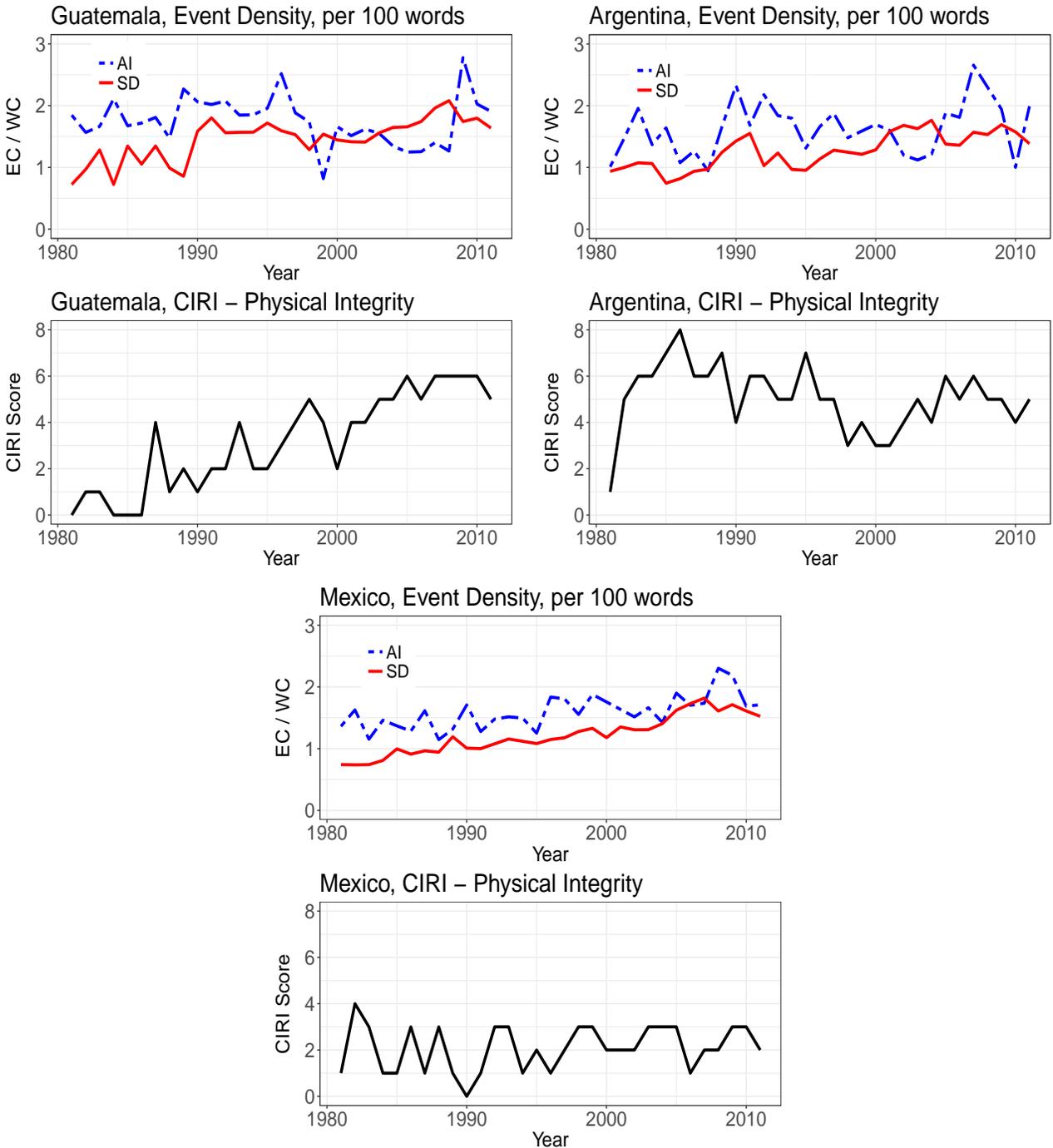


Figure 12: CIRI Physical Integrity Score vs. Event Density Over Time

<sup>20</sup>See Cingranelli and Richards (1999) for more information regarding the construction. See <http://www.humanrightsdata.com/p/data-documentation.html> for the documentation

Figure 13 graphs the proportion of events falling into each category for each country and organization. Each band of color represents the proportion of total events within a country-year report within a particular category. For more the raw frequency counts, see Figure 3 within the paper.

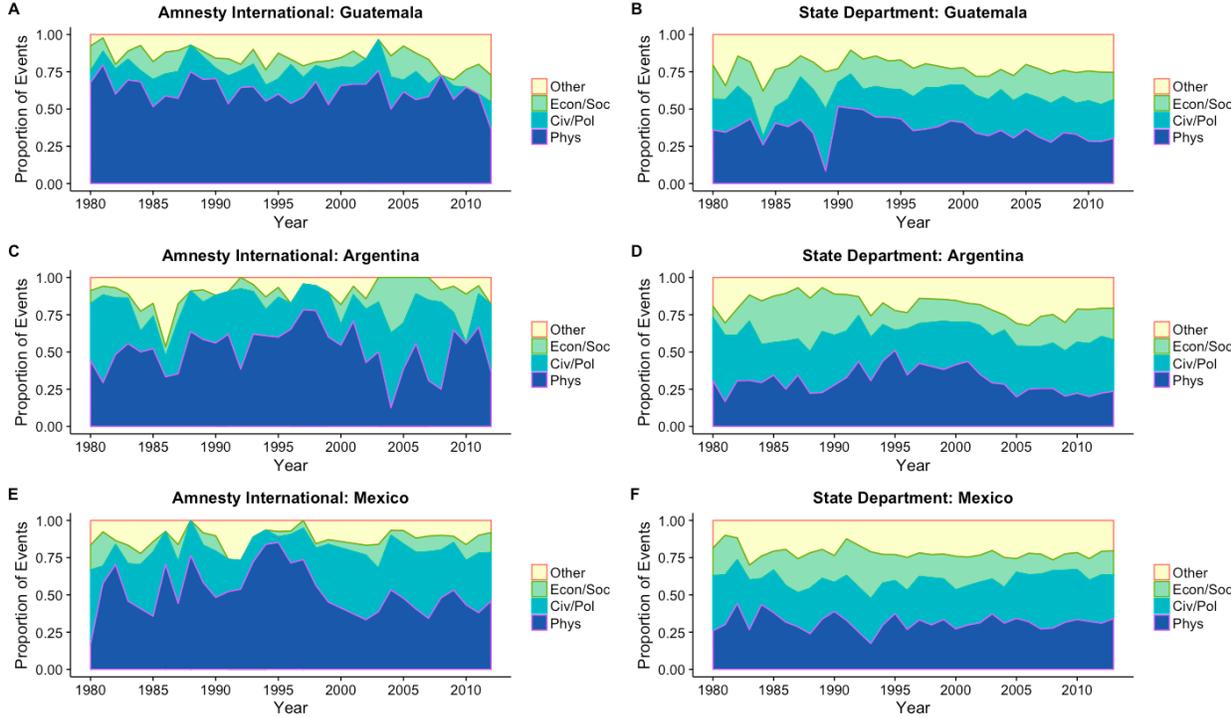


Figure 13: Verb Category Use as a Proportion of Events Over Time

### 8.3 Latent Semantic Analysis

Latent Semantic Analysis (LSA) is a theoretical and computational approach with origins in information retrieval (Bellegarda, 2007; Wild, 2016), and cognitive linguistics (Evangelopoulos, 2013). that leverages the relationship between words to tap into a document’s underlying latent semantic structure, understood here as the type of information within the reports. It can be used to understand the meaning of a particular document by computationally identifying the meaning of terms and mimicking human semantic processing. (Landauer et al., 1998) LSA creates a universe of semantic meaning by relating words within reports to each other. Documents are then plotted in this universe depending on the words they contain. We

then can retrieve how similar any document is to another by obtaining the cosine similarity or Pearson correlation coefficient, essentially providing us a measure of distance between any two documents in the latent semantic space.

Figure ?? plots the correlation between State Department and Amnesty country-reports for a given country-year suggesting evidence of global changes in reporting standards. This graph confirms what we already know about the process by which these reports are constructed from speaking with officials. There is intentional cross-referencing between organizations, reflected by the high level of correlation across all three countries.

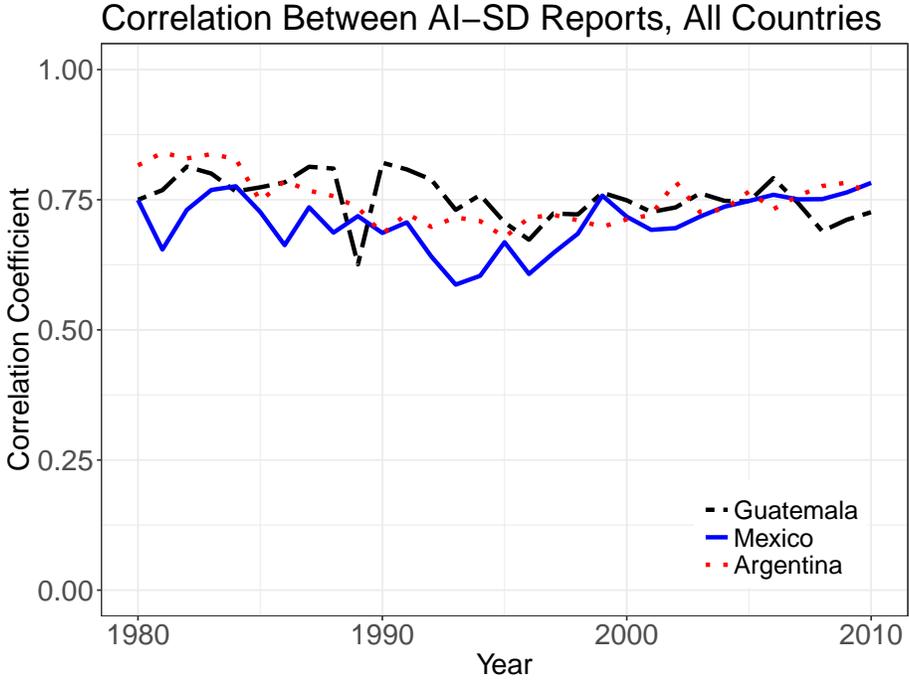


Figure 14: Correlation between Amnesty and State Department Country-Year Reports

A latent semantic space is defined by the relationship words have to each other and exists in a multivariate space of dimension  $k$ , where  $k$  is the number of terms in a corpus. While mapping terms individually does not capture their movement in relation to other terms, we can get a sense of what this space contains by drawing the top 10 highest loading terms in the document-term matrix for each document. This provides us with a rough, shorthand approximation of the position of the document in the latent semantic space.

Table 4: Top Loading Terms For Report, By Country

	1980		1990		2000		2010	
	AI	SD	AI	SD	AI	SD	AI	SD
<b>Argentina</b>	prison amnesti disappear intern right human report releas air bueno	govern right human polit report disappear argentina prison court organ	polic alleg investig militari report air bueno disappear right tortur	govern right polic argentina human law polit militari court labor	polic human right report investig prison arrest govern court provinc	govern right law right prison human report investig court judg	right provinc human former polic communiti indigen tortur prison peopl	govern report law court prison right provid labor air investig
<b>Mexico</b>	report arrest amnesti intern disappear mexico prison de peasant alleg	govern polit mexico percent mexican parti report claim prison pri	tortur right report human polic respons disappear arrest alleg commiss	govern mexico union polit right human feder labor polic law	right human report prison tortur de author mexico govern includ	govern right human report polic feder union prison mexico offic	right human polic investig report govern kill militari feder crimin	govern right law human report feder polic provid offic investig
<b>Guatemala</b>	kill intern govern disappear amnesti guatemala report forc peopl death	govern polit report right forc militari nation elect parti labor	right human report guatemala disappear govern militari amnesti intern death	right human guatemala govern forc union kill report polic guerrilla	right human guatemala govern report guatemalan death offici kill includ	govern right offic court report labor investig polic human union	right human kill death peopl women govern court offic crime	govern report law labor union kill right person public provid

## References

- Bagozzi, B. E. and Berliner, D. (2016). The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of US State Department Human Rights Reports. *Political Science Research and Methods*, page 1–17.
- Bellegarda, J. R. (2007). *Latent Semantic Mapping: Principles & Applications*. Morgan & Claypool Publishers.
- Caliendo, S. M., Gibney, M. P., and Payne, A. (1999). All the News that’s Fit to Print. New York Times Coverage of Human-Rights Violations. *The International Journal of Press/Politics*, 4(4).
- Clark, A. M. and Sikkink, K. (2013). Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures? *Human Rights Quarterly*, 35(3):539–568.
- Danzger, M. H. (1975). Validating conflict data. *American Sociological Review*, 40(5):570–584.
- Davenport, C. (2010). *Media Bias, Perspective, and State Repression: The Black Panther Party*. Cambridge University Press, New York City.
- Davenport, C. and Ball, P. (2002). Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995. *Journal of Conflict Resolution*, 46(3):427–450.
- David L. Cingranelli and Richards, David L. (1999). Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights. *International Studies Quarterly*, 43(2):407–18.
- de Neufville, J. I. (1986). Human rights reporting as a policy tool: An examination of the state department country reports. *Human Rights Quarterly*, 8(4):681–699.

- Escribà-Folch, A. (2013). Repression, political threats, and survival under autocracy. *International Political Science Review*, 34(5):543–560.
- Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(6):683–692.
- Fariss, C. J. (2014). Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability. *American Political Science Review*, 108(02):297–318.
- Fariss, C. J., Linder, F. J., Jones, Z. M., Crabtree, C. D., Biek, M. A., Ross, A.-S. M., Kaur, T., and Tsai, M. (2015). Human rights texts: Converting human rights primary source documents into data. *PLOS ONE*, 10(9):1–19.
- Hafner-Burton, E. and Ron, J. (2013). The Latin Bias: Regions, the Anglo-American Media, and Human Rights. *International Studies Quarterly*, 57(3):474–491.
- Kelley, J. (2007). Who Keeps International Commitments and Why? The International Criminal Court and Bilateral Nonsurrender Agreements. *The American Political Science Review*, 101(3):573–589.
- Kim, H. and Sikkink, K. (2010). Explaining the Deterrence Effect of Human Rights Prosecutions for Transitional Countries<sup>1</sup>. *International Studies Quarterly*, 54(4):939–963.
- Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3):259–284.
- Murdie, A. and Bhasin, T. (2011). Aiding and Abetting: Human Rights INGOs and Domestic Protest. *Journal of Conflict Resolution*, 55(2):163–191.
- Murdie, A. and Peksen, D. (2014). The Impact of Human Rights INGO Shaming on Humanitarian Interventions. *The Journal of Politics*, 76(1):215–228.

- Poe, S. C., Carey, S. C., and Vazquez, T. C. (2001). How are these pictures different? a quantitative comparison of the us state department and amnesty international human rights reports, 1976-1995. *Human Rights Quarterly*, 23(3):650–677.
- Powell, E. J. and Staton, J. K. (2009). Domestic Judicial Institutions and Human Rights Treaty Violation. *International Studies Quarterly*, 53(1):149–174.
- Richards, D. L. (2016). The Myth of Information Effects in Human Rights Data: Response to Ann Marie Clark and Kathryn Sikkink. *Human Rights Quarterly*, 38(2):477–492.
- Ron, J., Ramos, H., and Rodgers, K. (2005). Transnational information politics: Ngo human rights reporting, 1986–2000. *International Studies Quarterly*, 49(3):557–588.
- Schrodt, P. A. (2014a). TABARI: Textual Analysis by Augmented Replacement Instructions Version 0.8.4 Manual.
- Schrodt, P. A. (2014b). Tabari version 0.8.4. <http://eventdata.parusanalytics.com/software.dir/tabari.html>.
- Shattuck, J. (Oct. 11, 1993). Article 8: Violations of Women’s Human Rights. *US DEPARTMENT OF STATE DISPATCH*, 4(41).
- Vargas, T., Parkinson, S. E., and Bagozzi, B. E. (2016). Bleeding But Not Leading?: US Political Dynamics and Bias in International Event Data.
- Weber, R. (1990). *Basic content analysis*. Number no. 49 in Quantitative applications in the social sciences. Sage Publications.
- Weidmann, N. B. (2014). On the Accuracy of Media-based Conflict Event Data. *Journal of Conflict Resolution*, 59(6).
- Weidmann, N. B. (2016). A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science*, 60(1):206–218.

Wild, F. (2016). *Learning Analytics in R with SNA, LSA, and MPIA*. Springer, New York, NY, 1st ed. 2016 edition edition.

Wood, R. and Gibney, M. (2010). The political terror scale (pts): A re-introduction and a comparison to ciri. *Human Rights Quarterly*, 32(2):367–400.

Wood, R. M. (2008). “A Hand upon the Throat of the Nation”: Economic Sanctions and State Repression, 1976–2001. *International Studies Quarterly*, 52(3):489–513.