

What Bias? Changing Standards, Information Effects, and Human Rights Measurement

Peter Haschke* Daniel Arnon[†]

Tuesday 16th October, 2018

Abstract

Human rights measurement efforts have been confronted with concerns of bias practically since quantitative measurement efforts began in the late 1970s and early 1980s. By and large, attacks have focused on biases to the source materials on which these measurement efforts rely – namely the annual human rights reports produced by Amnesty International and the U.S. Department of State. We here take stock and seek to distinguish conceptually distinct types of bias that have been conflated in the past, yet plausibly affect human rights measures. In addition to revisiting reporting bias or organizational bias long identified in the literature, we also disentangle two types of bias that have been of concern more recently – bias attributable to changing standards and information effects. For each type of bias we identify, we provide an empirical implication as to the effect on human rights measures and importantly its spatial or temporal variation.

*Department of Political Science, University of North Carolina, Asheville, One University Heights, Asheville, NC 28805, email: phaschke@unca.edu

[†]Department of Political Science, Emory University, 201 Dowman Drive, Atlanta, GA 30322, email: darnon@emory.edu

1 Introduction

Quantitative measures of human rights violations/protection have become a crucial source of information for empirical work relating to human rights. The predominant standards based measures (e.g., the Political Terror Scale (PTS), the Cingranelli and Richards Physical Integrity Rights Index (CIRI)) have been cited hundreds of times in articles covering a wide range of topics such as domestic conflict processes, state repression, I/NGO behavior and influence on policy making. More recently, interest in the dis-aggregation of the human rights concept into various composite elements as well as methodological developments led to a proliferation of new indexes, including new measures of physical integrity violations, civil and political rights violations, torture or human trafficking. Another area of recent scholarship concerns itself with the validity of human rights measures and a lively debate has emerged among scholars producing and employing these indicators.

Among the main findings in quantitative human rights scholarship are the strong links between regime characteristics, democratic institutions, economic conditions and countries' human rights performances (Henderson 1982; Poe and Tate 1994; Davenport and Armstrong 2004; De Mesquita et al. 2005; Davenport 2007; Hill and Jones 2014; Haschke 2018). Since the third wave of democratization in the early 1990's, the world has seen many countries markedly improve their overall economic and political conditions, yet the human rights measures have indicated that the average level of respect for human rights globally has varied little if at all. This led scholars to question the validity of these indexes, as one would expect that overall economic and political improvements, and the work of governments and I/NGO's on improving human rights to lead to similar improvements of human rights conditions.

The main criticism rests on the notion that 1) the standards that are applied by

monitoring agencies that produce the human rights reports and which provide the source material from which standards-based human rights measures are produced, have changed, and that 2) the information environments in which monitoring agencies operate are vastly different today than 25 years ago (Clark and Sikkink 2013; Fariss 2014). Critics argue that because the reporting agencies compiling the reports have increasing budgets and better access to information over time, they are able to report more accurately on human rights violations. Assuming that producers of human rights measures continue to apply consistent *internal standards* for coding reports, more information about abuses and violations in human rights reports leads to the production of biased scores.

Any evidence of changing standards or changing information environments of course raises serious concerns about the ability of standards based measures to track human rights conditions over time. Potz-Nielsen, Ralston and Vargas (2018), for example, argue that because of these changing standards over time, standards based measures are entirely inappropriate for temporal comparisons and allow for only cross-sectional comparisons. Similarly, Eck and Fariss (2018) caution against cross-sectional comparison. Thus while questions and concerns as to validity and reliability have followed standards based measures of human rights for decades (see: Poe and Tate 1994; Poe, Carey and Vazquez 2001; Simmons 2009; Wood and Gibney 2010), this more recent scholarship on biased human rights measures spurred a different kind of controversy. Much of this controversy revolves around attempts to develop unbiased or less biased measures and specifically Fariss' (2014) finding that global human rights conditions have improved from the perspective of such de-biased measures.

Whereas some producers of standards based human rights measures (Cingranelli and Filippov 2017, 2018) are not convinced that efforts hoping to account and correct for biases actually produce more valid and trustworthy measures of human rights

conditions (see also: [Fariss 2017, 2018](#)), we here hope to contribute by reviewing and categorizing various sources of bias as well as their expected effect on standards based human rights measures.

We will begin our discussion with the two most prominent types of bias identified in recent scholarship – 1) bias attributed to changing monitoring and reporting standards and 2) bias associated with changing information environments. These two types of biases are at the heart of arguments advanced by [Clark and Sikkink \(2013\)](#) and [Fariss \(2014\)](#). We focus on these two because we believe both produce empirical predictions about patterns and trends that appear not to be borne out by the data. Following that discussion, we will identify a number of other sources of bias that have been identified in scholarship as well as their likely effects on human rights measures.

2 Types of Bias

A number of specific and distinct sources of bias have been identified in the literature. Below, we present a simple typology of these distinct biases that arguably affect standards based human rights measures (see: [Table 1](#)). The main dimension on which to distinguish biases concerns their spatial or temporal variability. Some bias affects all reports and as such all human rights scores in a given year equally. This bias is constant across reports but varies *temporally*. The biases identified by [Clark and Sikkink \(2013\)](#) and [Fariss \(2014\)](#) fall into this temporally varying category. Other biases affect only a certain subset of reports in a given year, leading to biased scores only for a subset of countries. This second type varies spatially. Whereas the former type makes temporal comparisons of human rights scores difficult (see for example: [Fariss 2014](#); [Pots-Nielsen, Ralston and Vargas 2018](#)), the latter likely complicates

Table 1: Typology of Bias

Bias	Variability	Hypothesized Effect	Location
Changing Standards	temporal	positive	monitors
Information Effects	temporal	positive	monitors
Reporting Bias I	spatial	positive or negative	monitors
Reporting Bias II	temporal	positive	monitors
Access Effects	spatial	negative?	countries
Transparency Effects	spatial	positive	countries
Coder Effects	temporal	positive	coders

cross-sectional comparison (see for example: [Eck and Fariss 2018](#)).

2.1 Changing Standards

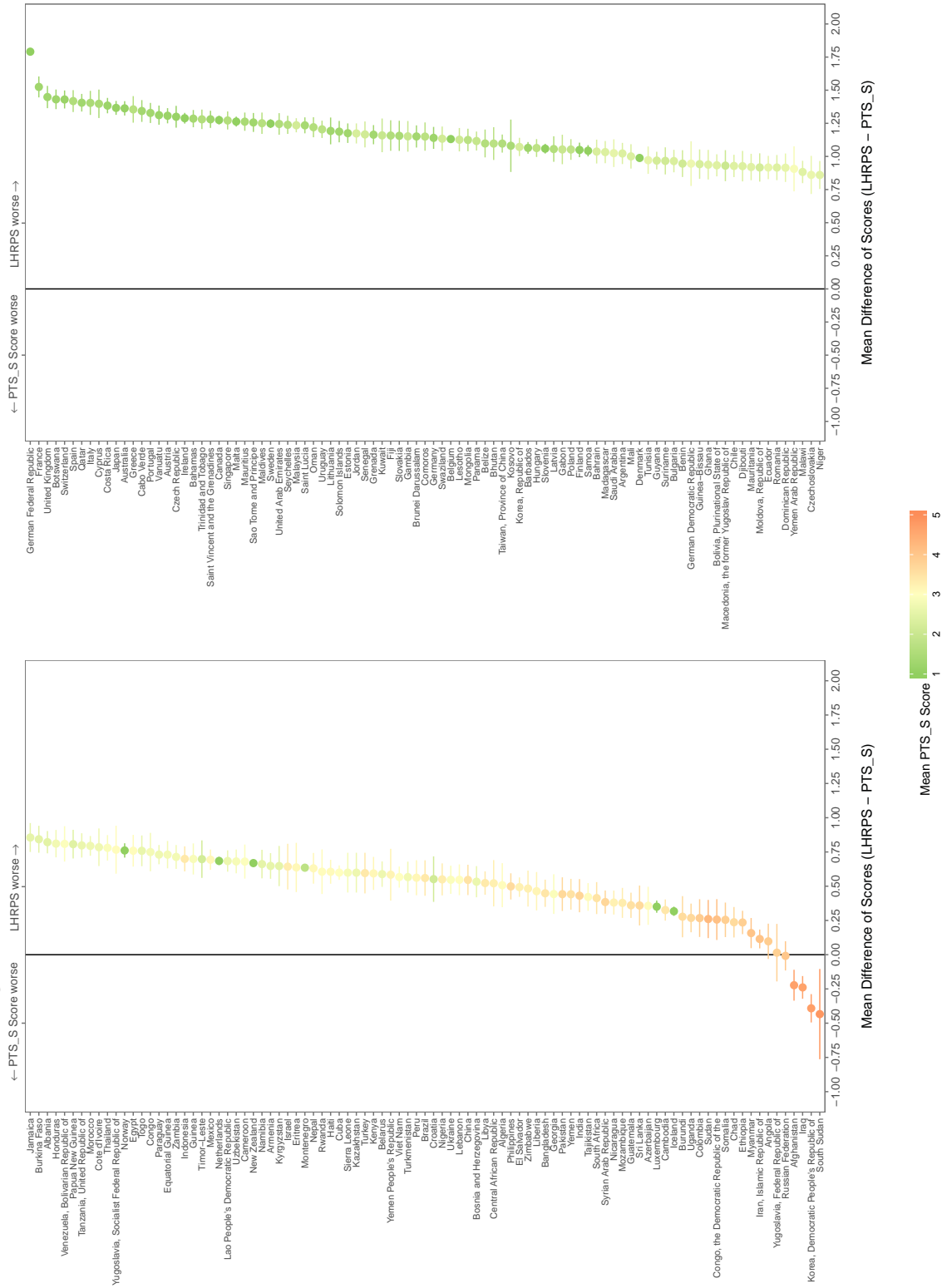
[Clark and Sikkink \(2013\)](#) and [Fariss \(2014, 2017, 2018\)](#) have argued that the standards monitoring agencies use to determine what constitutes abuse or a human rights violations, have changed dramatically since human rights reports began to be produced in the 1970s. The very “definition of what constitutes torture or state-sponsored killing has expanded over the years” and behaviors that were not considered human rights concerns in the past are considered violations today ([Clark and Sikkink 2013](#), 546). Early State Department reports, for instance, focused extensively on the most serious violations, such as extra-judicial killings and or forced disappearances. More recently, reports provide thorough detail of violations including excessive use of force, stealth torture, or stress and duress methods – practices that arguably would not have been included in earlier reports ([Clark and Sikkink 2013](#)). Increasingly strict standards and changing expectations used by monitoring agencies such as Amnesty International, the U.S. State Department or Human Rights Watch “mask real improvements to the level of respect for human rights” as these changing standards

translate into increasingly detailed and stringent reports (Fariss 2017, 239-40). This detail in the reports is then picked up by coders of human rights measures such as the PTS or CIRI and human rights scores will be biased because coders do not take into account what Fariss (2014) calls the “changing standards of accountability.”

While Clark and Sikink (2013) and especially Fariss (2014, 299) include and ultimately conflate conceptually distinct types of bias (i.e., bias attributed to changing information environments, access effects, and changing standards) into either “information effects” or “changing standards of accountability,” only what Fariss calls monitors’ “classification” strategies neatly fit into what we here call the *changing standard* category. We believe that Fariss’ (2014, 299) changes of the classification strategy or changes of monitors’ “subjective views of what constitutes a ‘good’ human rights record” are distinct from changes to monitoring capacity or changes to the level of access to information about human rights conditions. Bias attributable to changing standards is distinguishable in terms of both its effects on global averages as well as expectations as to its spatial and temporal variation.

Assuming constant human rights conditions globally, increasingly stringent and inclusive standards applied by monitoring agencies should produce the appearance of worsening global human rights conditions. Importantly, however, the effects of these changing standards should be constant across reports or countries and vary only temporally. Standards change over time but not from country to country. For example, if the State Department broadens its definition of torture it will likely do so across the board and for all countries. It is unlikely to apply different monitoring standards and definitions of what constitutes abuse to different countries. In other words, the State Department’s view of what constitutes a “good human rights record” should not change from country to country.

Figure 1: Mean Difference between PTS_S and Rescaled LHRPS



Note: Shown are the average differences between PTS_S and re-scaled LHRPS scores by country and 95% confidence intervals. Mean differences are colored. Red difference correspond to countries with on average poor human rights records (high PTS_S scores) green differences correspond to observations with on average good human rights conditions (low PTS_S scores).

Assuming that global human rights conditions have not changed over time, one would expect that average human rights conditions measured by the Political Terror Scale (PTS) or the Cingranelli Physical Integrity Rights (CIRI) index to diverge over time from average human rights measures that explicitly seek to account for temporal changes to standards applied by human rights monitors such as Fariss (2014)'s Latent Human Rights Protection Scores (LHRPS). This expectation is borne out and demonstrated by Fariss. Importantly, however, all PTS (or CIRI) scores in a given year should be inflated or biased equally as standards have changed over time but are consistently applied across countries. As such, differences between PTS and LHRPS should be uniform across countries and vary only over time.

However, a simple comparison of average differences between PTS and LHRPS across countries (as presented in Figure 1) shows that contrary to the above expectation, the differences between PTS and LHRPS are not constant across countries.¹ Rather PTS and LHRPS scores diverge (or converge) more or faster for some countries than for others.

What we find most surprising here is that these differences seem to be explained largely by human rights conditions. In fact, as shown in Table 2, human rights conditions as measured by PTS_S scores are a powerful predictor of the divergence (or convergence) of PTS_S and LHRPS scores. This cross-country variation over divergence (or convergence) of scores is inconsistent with the theoretical expectation that human rights standards have changed mostly over time. The PTS_S score should have no explanatory power if standards change over time but are applied consistently across countries in a given year. Additionally, the effects of time should be significantly more pronounced than estimated here. We discuss the alternative accounts

¹Note that LHRPS was re-scaled to range from 1 to 5 and inverted such that higher scores indicate worse human rights conditions rather than better human rights protection.

that could produce bias that varies across countries (e.g., *access* and *transparency effects*) below.

Table 2: Explaining PTS-LHRPS Divergence

	Model 1	Model 2	Model 3	Model 4
Intercept	54.490* (1.495)	2.021* (0.011)	42.920* (0.752)	37.444* (0.438)
PTS_S Score	—	-0.494* (0.004)	-0.471* (0.003)	-0.700* (0.003)
Year	-0.027* (0.001)	—	-0.021* (0.001)	-0.017* (0.001)
Adjusted R^2	0.174	0.694	0.793	0.937
Country Fixed Effects	no	no	no	yes

Note: Shown are coefficients from OLS regressions on the differences between PTS_S and rescaled LHRPS scores. $N = 6123$; standard errors in parentheses; * indicates significance at $p < 0.001$.

2.2 Information Effects

Akin to the changing standards account discussed above, [Keck and Sikkink \(1998\)](#), [Clark and Sikkink \(2013\)](#) and [Fariss \(2014\)](#) have identified an alternative source of a bias, which we here call an *information effect*.² Whereas in the changing standard account, monitors’ expectations and definitions of good human rights practices change over time, in this account it is the monitoring and research capacity that has changed over time. Monitors may not be classifying more acts as abuse but are able now to “look harder for abuse [and] look in more places” ([Fariss 2014](#), 297). Human rights monitoring capabilities of the State Department, for example, have grown exponentially since the 1970s when only a single State Department employee was responsible for the human rights reports ([Clark and Sikkink 2013](#)). By the mid-1990s, the State

²Also see [Richards \(2016\)](#) for a critical response.

Department's Human Right Bureau had ballooned to over 100 staff (Sikkink 2004). Additionally, more resources are available to embassy staff to report on conditions on the ground in order to evaluate and compile information from an increasing number of domestic human rights NGOs. In short, human rights monitors have access to increasing amounts of human rights information, and are better able (in terms of staffing and resources) to source and compile this information into their reports.

While bias attributable to changing expectations and standards is conceptually distinct from bias due to changes to an organization's monitoring and reporting capacity, we argue that their effects on human rights measures are likely quite similar. For instance, we expect increased staffing and growing budgets to the State Department's Bureau of Democracy, Human Rights and Labor to translate into improved monitoring of human rights conditions in all countries and as such vary mostly over time. Additional resources (but also budget cuts) are likely distributed more or less equally across the board and the additional 100 staff added to the State Department's human rights monitoring unit in the 1980s and 90s are unlikely to focus exclusively on human rights conditions in two or three countries. Similarly, human rights reports ought to feature more detail and increase in length equally or proportionally across countries (see: Haschke and Gibney 2018).³ Under constant human rights conditions, a growing monitoring capacity over time is then expected to lead to an appearance of worsening human rights conditions in the aggregate, as information effects are constant across countries but vary temporally.

³Haschke and Gibney (2018) show that the State Department's human rights reports have not grown appreciatively in length since 1999. Sections I of the reports in fact have become slightly shorter over the 17 year period.

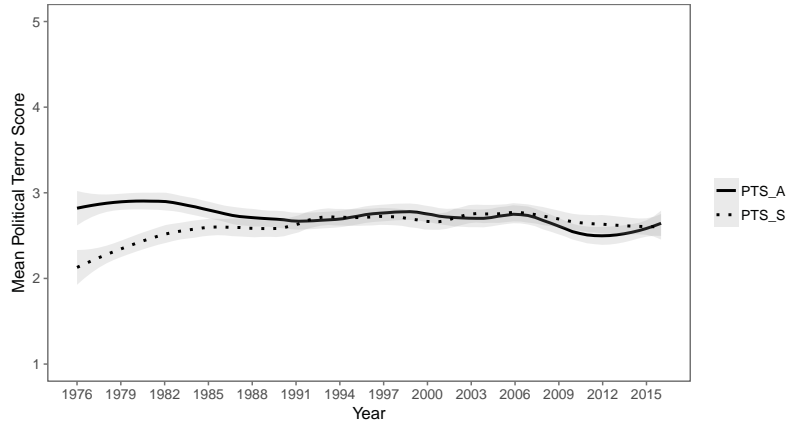
2.3 Reporting Bias I

Arguably, the first source of bias scholars identified could be called *reporting bias*.⁴ While standards based human rights measures such as the PTS or CIRI attempt to provide estimates of the state of physical integrity rights in a given country for a given year, producers of these measures do not and cannot claim that their scores represent the true state of a country’s human rights practices. Scores represent the state of human rights *as reported* by a given reporting organization. Amnesty International, for example, may report abuses which are not mentioned in the reports compiled by the U.S. State Department, while the State Department could focus on areas which did not receive the attention of Human Rights Watch. By and large then, human rights scores are a reflection of human rights records as seen and reported by the reporting organization. They may approximate the “truth,” but they likely contain some amount of subjectivity or bias depending on each organization’s mandate or agenda.

To provide an example: the U.S. State Department has long been accused of having strategically adjusted reporting standards for its annual reports produced in the 1970s and early to mid 1980s. According to a number of scholars (e.g., [Poe and Tate 1994](#); [Poe, Carey and Vazquez 2001](#); [Clark and Sikkink 2013](#); [Nieman and Ring 2015](#)), reports were allegedly biased in their content to make U.S. allies and U.S. foreign aid recipients appear more favorable. The bias originated with the monitoring agency itself and varied from country to country. The effect of this reporting bias is evident in [Figure 2](#), where global human rights conditions look much better from the perspective of human rights scores based on the U.S. State Department reports (PTS_S) compared to scores based on Amnesty International’s reports (PTS_A).

⁴[Clark and Sikkink \(2013\)](#) refer to this effect as “organizational bias.”

Figure 2: Difference between Annual PTS_S and PTS_A Averages



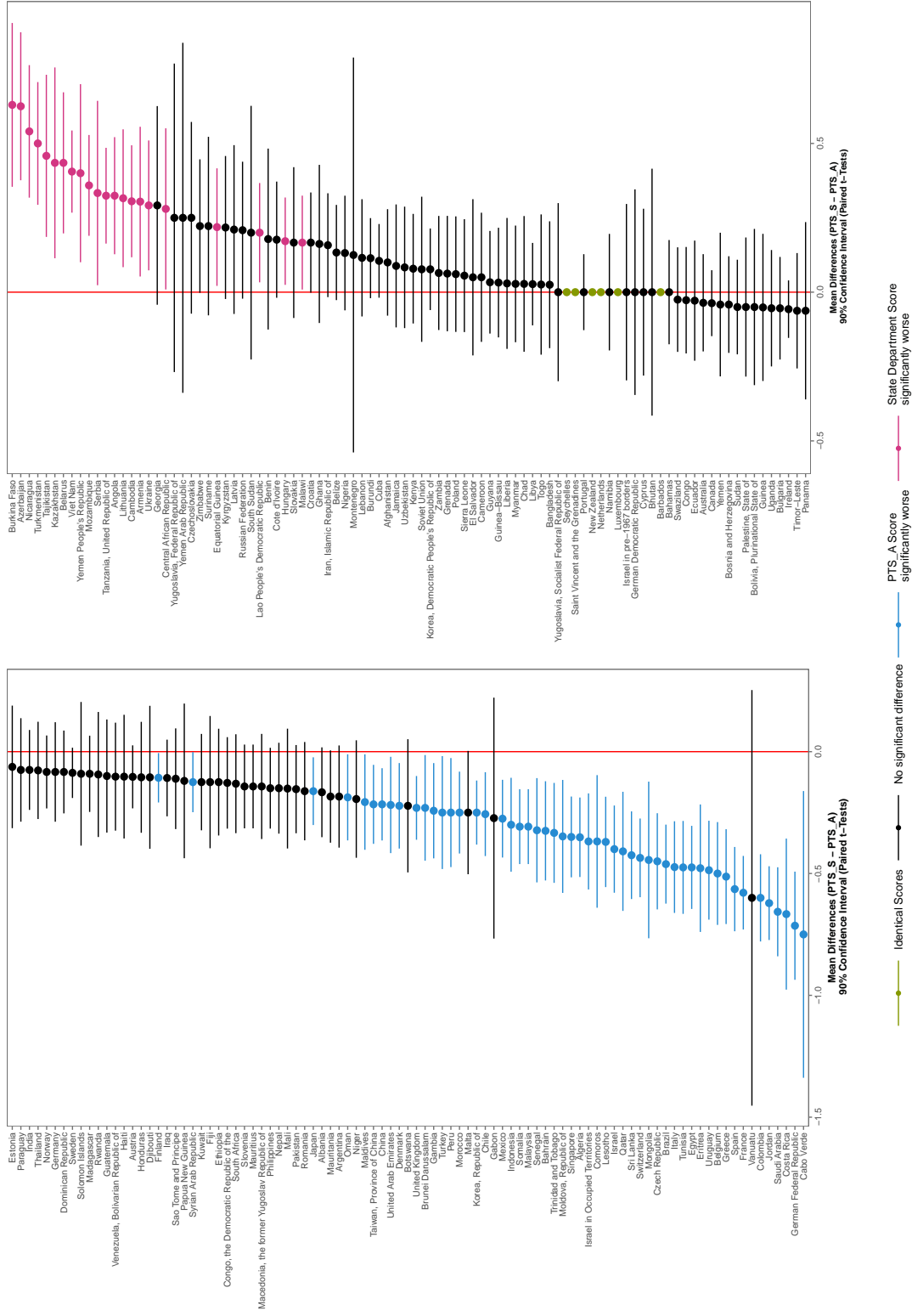
Note: Shown are smoothed annual average PTS_S (dotted) and PTS_A (solid) scores since 1976 for observations (countries) for which both PTS_S and PTS_A scores are available.

According to [Poe, Carey and Vazquez \(2001\)](#), the State Department reports ignored or at least downplayed physical integrity rights violations of U.S. allies, leading to biased PTS_S scores certainly in the 1970s. By the mid-1980s, average PTS_S scores converge on the PTS_A average. Arguably, however, human rights conditions did not improve from 1976 to 1986 but the State Department merely produced less biased reports and scores were no longer affected by *reporting bias*.

Assuming constant human rights conditions over time, monitoring bias can have either positive or negative effects on annual averages. If a monitoring agency begins exaggerating human rights abuses in some countries, global averages deteriorate despite unchanged conditions. Conversely, if monitoring agencies begin to ignore human rights violations in some countries, global human rights conditions will appear to be improving.

Importantly, reporting bias varies from country to country and does not affect all countries equally – like say bias attributed to *changing standards*. This variability of the bias across countries is highlighted in [Figure 3](#).

Figure 3: Difference between PTS_S and PTS_A Scores



Note: Shown are the mean differences between PTS_S and PTS_A scores by country and 90% confidence intervals. Countries with fewer than 4 observations have been omitted.

The figure presents the results of comparing two standards based human rights measures (PTS_S and PTS_A) for each country by means of a set of paired *t*-tests. Paired *t*-tests allow for a quick and dirty evaluation of ratings or biases of two raters (in our case two reporting agencies – Amnesty International and the State Department). Presented are the average differences of PTS_S and PTS_A scores by country, as well as 90 percent confidence intervals.⁵

Countries with dots on the right side of the red line (indicating no differences in scores), receive on average higher (worse) State Department scores. Countries with dots left of the red line, receive higher (worse) Amnesty scores. Magenta colored dots indicate that these differences are statistically discernible from zero (i.e., Amnesty scores are significantly lower or the State Department scores significantly higher). Blue colored dots indicate the opposite (i.e., Amnesty scores are significantly higher or the State Department scores significantly lower).

The results suggest that disagreement of scores is limited to a set of specific countries. The State Department’s reports are arguably more critical of former Eastern Bloc countries (e.g., Azerbaijan, Turkmenistan, Kazakhstan, Belarus, Georgia, and Russia). Other perhaps unsurprising examples include Vietnam, Mozambique, Angola, and Nicaragua. The State Department also appears to be significantly less critical of prominent U.S. allies such as West Germany, Colombia, Saudi Arabia, and Israel.

2.4 Reporting Bias II

Simmons (2009) has argued that non-governmental organizations such as Amnesty International have incentives to consistently report bad news even if states’ human

⁵Only countries with at least 4 complete observations (i.e., both PTS_S and PTS_A scores are available) were included.

rights records improve.⁶ If human rights records across the world improve sufficiently, Amnesty International’s ability to mobilize members and attract donations would arguably be eroded. In short, Amnesty International has an incentive to change its standards or to focus its attention to violations ignored in the past to remain relevant. Incentives to strategically adjust reporting standards in this way are likely going to affect all reports in a given year equally. As human right conditions improve globally, Amnesty International would feel pressured to make all reports appear to report bad news and one would not expect merely a handful of countries to provide an exaggerated or inflated account of human rights abuse. Of course, the State Department does not depend on member donations and should be free to report good news. The fact that PTS_A and PTS_S scores nevertheless differ more for some countries than for other (as shown in Figure 3) appears to be inconsistent with the expectation that all countries should be equally biased by Amnesty International’s alleged exaggeration of human rights abuse.

2.5 Access Effects

The information effects argument outlined above centers on changes to the information environment in which reports are created. As monitors’ budgets, staffing, and monitoring capacity increase, monitors are able to identify and gather more evidence of abuse that could be introduced into published human rights reports. Hill, Moore and Mukherjee (2013) find evidence that this increased capacity of the monitoring agencies is likely to reduce the bias from strategic selection of the reporting. Fewer resource constraints allow monitors to focus attention to more than just the most egregious violations. While increased capacity makes the monitoring process more consistent and less selective from the monitors’ perspective, the level of access

⁶See also: Hill, Moore and Mukherjee (2013).

to information that monitors encounter on the ground is still likely to vary across countries.

Human rights monitoring is often undermined by governments seeking “to hide, downplay, or dismiss information” (Clark and Sikkink 2013, 545). Human rights NGOs are banned, monitors harassed or expelled.⁷ Two otherwise similar monitoring teams (in terms of staffing and research capacity) sent to North and South Korea, respectively, will arguably encounter vastly different information environments. We call bias attributable to varying levels of access to countries *access effects*.

It is important to note here that *access effects* are closely related to *information effects* – both largely characterize the overall information environment. While the latter refer to the monitors’ ability or capacity to gather, research, and process information and evidence of human rights abuse, access effects refer to the restrictions and limits individual countries place on monitors to do their work. Information effects as characterized above originate with the monitoring agencies or organizations. Access effects originate with target countries. Unlike information effects, access effects are likely to vary across countries. The direction of the effect, however, is difficult to assess. Assuming de facto human rights conditions remain constant over time, increased restrictions to access to information should theoretically lead to a reduction in the number of reported violations of human rights and should give the appearance of improved human rights conditions.

Interestingly, when examining the PTS and LHRPS scores for a set of countries where monitors have been expelled, we see that scores remain unchanged.⁸ For ex-

⁷Recent examples of countries from which human rights monitors have been expelled include Burundi, Israel, Russia, Uzbekistan, and Venezuela. Not surprisingly, human rights practices in these countries are far from perfect and recent PTS_A or PTS_S scores are consistently 3 or higher.

⁸Considering that especially the U.S. State Department relies on NGO reporting in compiling its own reports, this expulsion of NGO monitors reduces access to relevant human rights information and should be reflected in the State Department’s report and the corresponding PTS_S score.

ample, after the expulsion of Human Rights Watch from Uzbekistan in 2011, we see no change in PTS_A and PTS_S scores. In fact, both Amnesty International’s and the State Department’s reports for Uzbekistan have been consistently coded as a “3” since 2006. We also observe little if any change in the LHRPS score, which at best shows a slight improvement in protection 2006-2010, but a slight decrease in protection since then. We should note that this anecdotal evidence is merely suggestive and cannot account for other changing factors.

2.6 Transparency Effects

Related to access effects discussed above, another subcategory of information effects are *transparency effects*.⁹ Whereas some countries choose to intentionally limit access to information about human rights practices, others aim to improve transparency and access. Sweden for instance is found to differ dramatically in terms of its human rights documentation processes from countries such as Benin, Laos, or Belarus – countries that frequently receive similar or even identical scores on human rights measures as Sweden (Eck and Fariss 2018). The process of documenting human rights abuses for Sweden differs insofar as Sweden has developed a centralized databases to register all allegations of police violence and misconduct, most other countries have not (596). When producing their annual human rights reports, monitoring agencies such as the U.S. Department of State or Amnesty International consult these publicly accessible databases and information about allegations is incorporated into country reports.¹⁰ In countries without such transparency efforts, information concerning human rights conditions will be harder or even impossible to obtain, especially if officials seek to restrict the flow of information and hinder efforts to corroborate allegations of abuse.

⁹It is important to note that compared to the information effects bias, the sources of transparency effects and access effects is with countries compared to the monitoring agencies.

¹⁰See Eck and Fariss (2018) for detail.

As a consequence, a country report for Sweden will provide more documentation and detail of human rights violations than the report for Belarus – despite the fact that human rights conditions in the latter country are arguably worse.

Akin to [Conrad, Haglund and Moore \(2013\)](#), [Eck and Fariss \(2018\)](#) argue that any existing and future human rights measurement projects have to account for this variation in institutional transparency. While transparency varies over time, temporal variation is likely dwarfed by variation across countries and its effect on global averages is likely to be negative. Assuming constant human rights conditions, transparency efforts by some countries produce an inflation of relevant information about abuses and will lead to the appearance of worsening human rights conditions. As shown by [Eck and Fariss \(2018\)](#) this makes cross-sectional comparison problematic.

2.7 Coder Effects

Another potential source of bias – one that has largely been ignored in the literature – are the coders of standards based human rights measures who assign scores to individual human rights reports. What we call *coder effects* here refers to the bias that is introduced through inconsistent application of coding rules (as outlined in code books). This inconsistent application of the coding rules could be a function of turnover among coders as well as coders' changing interpretations of those coding rules. [Clark and Sikkink \(2013\)](#), [Fariss \(2014\)](#), and [Eck and Fariss \(2018\)](#) consider possible biases introduced in the process of compiling the report itself (i.e., at the level of the monitor) or during the publicizing of human rights violation (i.e., at the country level), but they ignore the possibility of the introduction of bias during the scoring of the reports (i.e., at the level of the coder).

Consider the coding schemes used by the CIRI and PTS projects. Both seek to

produce a general assessment of human rights violations as they are described in the human rights reports by the U.S. State Department or Amnesty International and Human Rights Watch.¹¹ Importantly, both projects ask coders to provide consistent and objective judgments of relevant human rights information contained in reports. Wood and Gibney (2010) nevertheless acknowledge that the PTS “relies heavily on subjective coding to generate a country’s score” (374). CIRI tries more explicitly to eliminate subjective coder judgments from the coding process. Coders here are provided specific instructions and precise numeric thresholds of reported violations to score reports when this information seems more reliable than the general characterization of human rights abuse such as for example “gross” or “widespread.”

In each scheme, coders ultimately still face choices that require at least some subjective judgment. Even when these choices are made transparently and are reconciled through an inter-coder arbitration, they still require that coders are able to ignore their own personal biases, to put aside what they might already know about a country and its human rights practices, and that assessments of reports will be consistent temporally and spatially. For example coders are asked to approach a report for the Democratic Republic of the Congo the same way they might approach a report for Liechtenstein. Similarly, coders are expected to evaluate reports the same way year after year.

Both PTS and CIRI attempt to train coders well. They aspire to systematically translate a given report’s content into numeric values, and both report high inter-coder reliability. While we are not challenging the coding process employed by the CIRI or PTS teams, we are questioning an underlying assumption about the location or origin of bias and ask whether standards applied by coders of human rights reports have changed over time. It is possible that the very changes in international norms and

¹¹Reports by the latter two are used only in the PTS measurement effort.

standards that influence the reporting agency’s report production process also affect the coders of the reports. In other words, perhaps the subjective assessments coders make when coding reports have changed. While the coding rules have remained constant, coders have not and their reading and interpretation of the coding rules could very well have changed over time.¹² As such human rights scores could be biased insofar as they capture coder specific changes. The very processes of developing international norms and changing human rights expectations, as outlined by [Clark and Sikink \(2013\)](#) or [Fariss \(2014\)](#), may affect the coding process as well.

Assuming constant global human rights condition, these *coder effects* will produce the appearance of worsening human rights practices. Similar to the bias attributable to changing standards, the effects will be constant across reports or countries and vary primarily over time. As coders change or coders approach reports differently year after year, bias will accumulate making temporal comparison challenging.¹³

2.8 Other Biases

Another source of bias in human rights global averages is the number of countries in the international system. The creation of new states in the international system, influences these averages by changing the number of observations in the sample. Consider for example the Soviet Union, which received an average PTS_S score of 2.85 until its collapse in 1991, including a 4 in its last year. In the following year, the PTS team began producing scores for each of the Soviet Union’s successor states (e.g., Azerbaijan, Tajikistan, Ukraine) separately, adding 14 new countries to the global average. The average PTS_S score for these 14 former Soviet Republics was 2.2. The

¹²Coders – often undergraduates or graduate students – tend to stay with the project for multiple years. However, there is some turnover and only very few coders have remained active over the entire 40-year period the PTS has been produced.

¹³An experiment to assess this hypothesis and identify potential *coder effects* in the PTS measurement project is under way.

addition of these 14 shifted their weight from being a single observation among 155 countries for which a score was given in 1991 (0.6 percent of the sample), to being 15 observations out of 166 countries (almost 9 percent of the sample). The global mean in 1992 was 2.46, but if we replace these new observations only with the score for the newly created Russian Federation, the global mean increases to 2.54, implying that the additional countries overall reduced the global average and generated the appearance of improving human rights conditions globally.

In 1980, the number of countries which received PTS_S scores was 144. In 1990, the sample size increases to 153, and by 2000 increases to 179. Just between 1990 and 1993, the number of reports which were coded increased by 21. The countries which were added by 1993 had lower average PTS scores than the sample of countries that had existed before 1993. As such, the addition of new countries decreased the global mean by 0.1 from 2.5 to 2.4 even though human rights conditions arguably had not changed. Importantly, some of the new countries also had a poor human rights records following independence. Bosnia and Herzegovina and Croatia are examples. Both countries saw improvements of their scores only after some time, dropping from an initial score of “5” to a score of “1,” respectively.

In addition to changes to the state system – with new countries emerging, others disappearing – the number of reports that are produced by monitoring agencies has varied independently (increasing over time). Similarly, the number of reports that are actually coded by the PTS and CIRI projects has changed as well. The PTS measurement effort only began in 2014 to code small island nations and micro states, such as Palau, Saint Kitts and Nevis and Nauru, or San Marino and Liechtenstein. The coding of reports for these states is likely responsible for the appearance of improving human rights conditions since 2014.

3 Conclusion

We have identified a number of conceptually distinct processes that could bias human rights measurement efforts. We believe that these processes are not only distinct in terms of the directionality of their effects but distinct in terms of their variation across countries, human rights reports, and over time. We distinguished biases at the monitor-level (e.g., monitoring effects, changing standards, information effects) from those originating with countries (e.g., access and transparency effects), and finally those that originate with the producers of human rights measures (i.e., coder effects). We characterized some bias as largely varying temporally whereas other processes make cross-country comparison problematic. Importantly, we argue that the bias processes identified in recent scholarship are theoretically expected to bias human rights measures mostly over time – affecting all human rights scores in a given year equally. Finally, we provided some suggestive evidence that is inconsistent with theoretical accounts of changing standards and information effects but in line with access or transparency effects.

References

- Cingranelli, David L., David L. Richards and K. Chad Clay. 2014. "The Cingranelli-Richards (CIRI) Human Rights Dataset." *CIRI Human Rights Data Project* .
URL: <http://www.humanrightsdata.org> (Retrieved August 2018)
- Cingranelli, David and Mikhail Filippov. 2017. "Problems of Model Specification and Improper Data Extrapolation." *British Journal of Political Science* pp. 1–2.
- Cingranelli, David and Mikhail Filippov. 2018. "Are Human Rights Practices Improving?" *American Political Science Review* . Letter.
- Clark, Ann Marie and Kathryn Sikkink. 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.
- Conrad, Courtenay R., Jillienne Haglund and Will H. Moore. 2013. "Disaggregating Torture Allegations: Introducing the Ill-Treatment and Torture (ITT) Country-Year Data." *International Studies Perspectives* 14:199–220.
- Davenport, Christian. 2007. *State Repression and the Democratic Peace*. Cambridge, MA: Cambridge University Press.
- Davenport, Christian and David A. II Armstrong. 2004. "Democracy and the Violation of Human Rights: A Statistical Analysis from 1976 to 1996." *American Journal of Political Science* 48(3):538–554.
- De Mesquita, Bruce Bueno, Feryal Marie Cherif, George W. Downs and Alastair Smith. 2005. "Thinking Inside the Box: A Closer Look at Democracy and Human Rights." *International Studies Quarterly* 49(3):439–458.
- Eck, Kristine and Christopher J. Fariss. 2018. "Ill Treatment and Torture in Sweden: A Critique of Cross-Case Comparison." *Human Rights Quarterly* 40:591–604.
- Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108(2):297–318.

- Fariss, Christopher J. 2017. "The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance." *British Journal of Political Science* pp. 1–33.
- Fariss, Christopher J. 2018. "Yes, Human Rights Practices Are Improving Over Time." *American Political Science Review* . Letter.
- Gibney, Mark, Linda Cornett, Reed Wood, Peter Haschke, Daniel Arnon and Attilio Pisanò. 2018. "The Political Terror Scale 1976-2017." *The Political Terror Scale* .
URL: <http://www.politicalterroryscale.org> (Retrieved Aug. 2018)
- Haschke, Peter. 2018. *Human Rights in Democracies*. New York, NY: Routledge.
- Haschke, Peter and Mark Gibney. 2018. Are Global Human Rights Conditions Static or Dynamic? In *Peace and Conflict 2017*, ed. David A. Backer, Ravi Bhavnani and Paul K. Huth. New York, NY: Routledge chapter 10, pp. 88–100.
- Henderson, Conway. 1982. "Military Regimes and Rights in Developing Countries: A Comparative Perspective." *Human Rights Quarterly* 4:110–123.
- Hill, Daniel W. Jr, Will H. Moore and Bumba Mukherjee. 2013. "Information Politics Versus Organizational Incentives: When Are Amnesty International's "Naming and Shaming" Reports Biased?" *International Studies Quarterly* pp. 1–14.
- Hill, Daniel W. and Zachary M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(03):661–687.
- Keck, Margaret E. and Kathryn Sikkink. 1998. *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.
- Nieman, Mark David and Jonathan J. Ring. 2015. "The Construction of Human Rights: Accounting for Systematic Bias in Common Human Rights Measures." *European Political Science* 14(4):473–495.
- Poe, Steven C. and Neal C. Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88(4):853–872.

- Poe, Steven C., Sabine C. Carey and Tanya C. Vazquez. 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* 23:650–677.
- Potz-Nielsen, Carly, Robert Ralston and Thomas R. Vargas. 2018. "Recording Abuse: How the Editing Process Shapes our Understanding of Human Rights Abuses." Working Paper.
- Richards, David 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.
- Sikkink, Kathryn. 2004. *Mixed Signals: U.S. Human Rights Policy and Latin America*. Ithica, NY: Cornell University Press.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge, MA: Cambridge University Press.
- Wood, Reed M. and Mark Gibney. 2010. "The Political Terror Scale (PTS): A Re-introduction and a Comparison to CIRI." *Human Rights Quarterly* 32(2):367–400.