

# The Political Terror Scale (PTS): A Re-introduction and a Comparison to CIRI

Reed M. Wood\*

Mark Gibney\*\*

## Abstract

Despite the frequency with which scholars have utilized the Political Terror Scale (PTS), a surprising number of questions remain regarding the origins of the scale, the coding scheme it employs, and its conceptualization of "state terror." This research note attempts to clarify these issues. We also take this opportunity to compare the PTS with the Cingranelli and Richards Human Rights Data Project (CIRI). Although the PTS and CIRI are coded from the same source material and capture the same class of human rights violations, we observe some important differences between the two that we believe may be of interest to scholars in the quantitative human rights community. First, we believe that the CIRI claims a level of precision that is not possible given the source data from which both datasets are coded. We believe that the PTS offers a transparent coding system that recognizes the inherent limitations in measuring abuses of physical integrity rights. Second, we argue that the CIRI's method of summing across abuse types leads to some inappropriate categorizations. For instance, the absence of one type of abuse prevents a state from being coded into a more repressive overall category regardless of the levels of other types of abuse. Lastly, the PTS accounts for the "range" of violence committed by the state—in short, what segments of the population are targeted. We believe that range is an important dimension to consider in measuring human rights and one to which CIRI does not attend.

## I. Introduction

Beginning in the early 1980s, scholars began to systematically investigate state-sponsored terror and violations of human rights.<sup>1</sup> In addition to significant contributions to the theoretical understanding of patterns of state-sponsored violence, scholars of this research agenda constructed the first cross-national measures of state violations of citizens' basic human rights.<sup>2</sup> The Political Terror Scale (PTS) was among the first quantitative datasets on state respect for human rights, and during the past quarter century it has been the most commonly used indicator of state violations of citizens' physical integrity rights.<sup>3</sup>

Despite the frequency with which scholars of political violence, human rights, and state

repression have utilized the PTS, a surprising number of questions remain regarding its origins, the coding scheme it employs, and its conceptualization of "state terror." This research note attempts to clarify these issues. We also take this opportunity to compare the PTS with the Cingranelli and Richards Human Rights Data Project (CIRI).<sup>4</sup> Although the PTS and CIRI are coded from the same source material and capture the same class of human rights violations, we observe some important differences between the two that we believe may be of interest to scholars in the quantitative human rights community. It is not our intention to promote the PTS over CIRI. While we disagree with the CIRI's creators on a number of conceptual and measurement issues, we do believe that the dataset has made an important contribution to this field.

## II. What's in a Name?

The PTS is a standards-based human rights data set first created by a group of colleagues at Purdue University in the early 1980s.<sup>5</sup> The name itself has engendered a fair amount of confusion. At times it has been called (incorrectly) the Purdue Terror Scale, no doubt because for some period of time the data were collected and disseminated by scholars who were then teaching at Purdue University. It has also been called (again, incorrectly) the Poe-Tate Scale. Steven Poe and Neal Tate did help with some of the coding for a few years, but what really seems to be behind this confusion are the initials of the last names of these two scholars (PT), combined with the outstanding work they have done using data from the Political Terror Scale.<sup>6</sup> However, there has never been a Purdue Terror Scale or a Poe-Tate Scale—rather, there has only been the Political Terror Scale. The original version of the PTS coded fifty-nine countries

for the years 1976-1983.<sup>7</sup> In 1984, Mark Gibney took over as PTS manager and he has served in this capacity ever since. The PTS has been expanded to more than 180 countries and it provides data on human rights practices for more than three decades (1976-2008).

Another apparent source of confusion is the use of the term “terror.” The PTS measures “state terror”: violations of physical or personal integrity rights carried out by a state (or its agents). This category of human rights violations will be familiar to scholars of state repression and political violence and includes abuses such as extrajudicial killing, torture or similar physical abuse, disappearances, and political imprisonment. This, of course, is different from the meaning that “terror” has now taken on, particularly after the events of 11 September 2001.

What should also be noted is that the PTS measures actual violations of physical integrity rights more than it is measuring general political repression. In fact, there will be instances where a government is so repressive that there might be relatively few acts of political violence. If a state is truly efficient in its use of coercive repression against its citizens, the repression from an earlier period should retain some "half-life" that continues to deter challenges to the regime in subsequent periods.<sup>8</sup> For example, the former Soviet Union received a PTS score of 2 or 3 most years during the late 1970s and early 1980s—a score that, without question, does not reflect the overall level of political repression and social control employed by the totalitarian regime during that period. Yet, the fact that the USSR had engaged in massive organized violence against its population during earlier periods, coupled with the ability of the state to monitor and police its population, meant that it did not need to resort to high levels of explicit violence during the 1970s and 1980s in order to keep its population in check.

The PTS focuses on state behavior. As such, domestic (family) or societal (mob, clan) violence, which, admittedly, are of epidemic proportions in many countries, are not included in a

country's annual score. Female genital mutilation and similar practices are also not measured by the PTS. Once again, the rationale is that this violence is carried out by private actors, which we try to differentiate (to the extent that it is feasible) from violence carried out by political actors such as the state and its agents (e.g., paramilitaries and death squads). That said, the PTS does not code violence ascribed to the actions of insurgent groups, criminal syndicates, gangs, or similar non-state actors whose motives may be political. For example, the violence committed recently by criminal drug syndicates in northern Mexico will not be considered in the state's score; however, extrajudicial execution and abuses in custody of detainees committed by security forces in the prosecution of the war on these organizations will be considered.

Abuse or deaths of detainees while in custody is one of the components of a state's PTS score that presents a number of potential complications. While the PTS does consider violence carried out by prison officials—torture is the most common example—the fact that a country's prison conditions are “harsh” or “life threatening” is not considered in determining a state's PTS score. A particularly troubling issue over the years has been the use of state sanctioned executions. The PTS focuses on state-sponsored killings that take place outside of the normal judicial setting. These "extrajudicial" executions or killings include death squad killings of political enemies, unlawful use of lethal force by police forces (e.g., shooting unarmed suspects), intentional killing of civilians by security forces during combat, and other arbitrary deprivation of life. However, the PTS does not include state executions that occur after trials that conform in at least some minimal way to international standards. As with many issues, however, what constitutes a legitimate “legal” execution and what constitutes an “extrajudicial” killing is difficult to determine. As a general rule, the PTS will code summary executions or those that take place outside the context of a legal proceeding as illegitimate executions and exclude those

that take place after legal proceedings.

Notwithstanding this focus on the state, there are occasions when it is not clear if it is “the state” that is the actor directly responsible for a given abuse. This is particularly a problem when paramilitary organizations or local militias engage in (as they often do) a significant amount of violence. It is not usually easy to distinguish the extent of government involvement in the activities of paramilitaries groups—this is often the reason governments choose to allow these organizations to operate. It is therefore incumbent upon coders to make these decisions. The reports from which we generate the scores often provide some insight into the level of involvement, and we thus rely on this information to make these decisions. Colombia is a case in point. Gross human rights violations by official state security forces have diminished markedly since the early 2000s. At the same time, violence committed by paramilitary organizations did not show a similar decline. Since the late 1990s paramilitary organizations such as the Autodefensas Unidas de Colombia (AUC) have been the most violent actors in Colombia.<sup>9</sup> While the state has made significant public efforts to curtail paramilitary violence in recent years, the annual reports of both Amnesty International (AI) and the US Department of State (USDS) provide evidence that key elements within the government and military provide support to or are complicit in the violence attributed to armed groups such as the AUC. As a result, the scores provided for Colombia are intended to reflect the close relationship between the government and these organizations. Other examples have included cases in which off duty police officers form social cleansing groups that target homosexuals, drug dealers, prostitutes, and other “social undesirables” or when security personnel have provided material support for the abduction of political figures or journalists or when they are complicit in kidnappings for profit.

### III. Coding the PTS

The data for the coding comes from two annual sources: the U.S. Department of State *Country Reports on Human Rights Practices* and the Amnesty International *Annual Report*.<sup>10</sup> In the construction of this index for each year, each report is scaled as if the information in the reports was accurate and complete. Thus, any biases found in the annual reports of the two organizations should be evident in the indices.<sup>11</sup> Each country in each report is coded by at least two senior coders,<sup>12</sup> and this is supplemented by the work of several students. Inter-coder reliability over the past five years has been over 0.85 among the principal coders. When disagreements do occur, principal coders review the disputed cases without referencing their previous decisions. If the principal coders still disagree, a third coder will be brought in. This seldom occurs in practice.<sup>13</sup>

The PTS uses a five point coding scheme that was adopted from a “political terror” scale published by Freedom House in its 1980 yearbook.

Level 1: Countries . . . under a secure rule of law, people are not imprisoned for their views, and torture is rare or exceptional. . . . Political murders are extremely rare . . . .

Level 2: There is a limited amount of imprisonment for nonviolent political activity. However, a few persons are affected; torture and beating are exceptional. . . . Political murder is rare . . . .

Level 3: There is extensive political imprisonment . . . . Execution or other political murders and brutality may be common. Unlimited detention, with or without trial, for political views is accepted . . . .

Level 4: The practices of Level 3 are expanded to larger numbers. Murders, disappearances, and torture are part of life . . . . In spite of its generality, on this level terror affects primarily those who interest themselves in politics or ideas.

Level 5: The terrors of Level 4 have been extended to the whole population . . . . The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals.<sup>14</sup>

While the descriptions of the coding categories are by now well known to most PTS users, fewer are likely aware of the underlying conceptual dimensions along which these "levels" of abuse are constructed. The underlying conceptual intuition of the PTS is that state violence can be assessed along three dimensions: scope, intensity and range.<sup>15</sup> In brief, *scope* refers to the type of violence being carried out by the state (imprisonment, torture, killing, etc.).<sup>16</sup> *Intensity* refers to the frequency with which the state employs a given type of abuse, more basically, the instances of a given type of abuse that are observed over a given period of time. *Range* is the portion of the population targeted for abuse. Readers should note that while intensity refers to the number of abuses committed by the government, range is intended to capture what segment(s) of society the government targets. This might also be seen as the selectivity of the violence.<sup>17</sup>

As the above discussion suggests, the PTS relies heavily on inter-subjective coding to generate a country's score, largely because we believe that the contextual factors found in the reports effectively prohibit purely objective coding criteria. It is critically important that coders understand both the intent and the underlying logic of the PTS and that they take care to avoid common pitfalls that come with subjective coding. Coders are instructed to ignore their own feelings and biases and to make every effort to assign a score that reflects what appears in the human rights report itself. In brief, they are disallowed from injecting their own "knowledge" of a case and to code based only on what the report actually says. Coders are also instructed to give countries the benefit of any doubt by providing a lower (or better) score when a report seems to fall somewhere between two numbers. Thus, if a coder believes that a country's score is between 2 and 3, s/he should code that country a 2. Perhaps the most elementary rule of all is that the coders need to see the PTS as representing a continuum of human rights practices. All else being equal, a level 3 country will experience higher levels of human rights violations than countries

coded at either level 2 or level 1, but lower levels of political violence than states coded at level 4 or level 5. Coders must also understand that the scores assigned to countries reflect similar groupings. All of the states within a given category should be characterized by similar patterns of abuses. In brief, all category 1 states will appear more or less similar in terms of abuse while all category 5 states will appear similar to one another in this respect.

We are not afraid to admit that commonsense is enormously important to this enterprise, and what we try to avoid is a mechanical application of the coding scale that arrives at absurd results. For example, level 5 speaks about political violence being extended to the entire population of a country. However, if gross and systematic human rights violations are being directed at a specific group—Rwandan Tutsis in 1994 or ethnic Albanians in Kosovo in 1999—this should not result in a lower (better) score simply because not all elements of a country's population are equally at risk. While this might seem to contradict the above discussion of range, two points are worth noting. First, the target population may be delimited along ethnic or religious lines; however, frequently the state's use of violence against such sub-sections of the society falls particularly high on the two other dimensions that underlie the logic of the PTS. In the above examples, while an ethnic minority was the immediate target of the state-sanctioned violence, the government employed violence with such frequency and brutality that both received scores of 5.<sup>18</sup> Second, and perhaps more important, range is intended to differentiate among targets of abuse based on their observable actions, behaviors, or associations. That is, all else equal, the PTS would assign a lower score to a state in which hundreds of political activists, labor leaders, and protesting students were killed compared to a state in which the same number of apolitical peasants were summarily executed. This is, distinction is most apparent in the scoring rules for categories 4 and 5.

#### IV. Comparing the PTS with the CIRI

Recently, scholars from SUNY Binghamton introduced a new human rights index known as the CIRI.<sup>19</sup> While we welcome the CIRI's valuable contribution to the field of quantitative human rights studies, we believe that users should be aware of the key differences between the datasets in order to evaluate which data are most appropriate for their particular needs.<sup>20</sup> In this section we seek to address these in some detail.

Like the PTS, the CIRI measures state-sponsored violations of the subset of human rights known as physical integrity rights.<sup>21</sup> In addition, both measures derive their respective categorical scores by subjectively coding the same source data. It is therefore unsurprising that the two measures are rather highly correlated—approximately 0.73 for the USDS-derived PTS scores and 0.65 for AI scores.<sup>22</sup> That both the PTS and CIRI focus on the same types of violence, code from the same descriptive data, and correlate reasonably highly suggests a high degree of similarities between the datasets.

Despite the similarities, the two datasets exhibit a few notable differences. First, CIRI explicitly disaggregates physical integrity violations into several (though not necessarily all) of its component parts: disappearances, killing, torture, and imprisonment. Second, it attempts to establish more precise threshold values for each category of intensity. Both represent potential improvements over the coding system and structure used by the PTS. Third, the datasets differ in their underlying logic: While the PTS relies on the three conceptual components discussed above and presents a standards-based ranking of government abuses, the CIRI explicitly assesses the frequency and types of government abuse practices. The two scores are therefore likely to paint somewhat different pictures of violence while still remaining highly correlated overall. We discuss these differences in turn and then offer what we believe is the first systematic

comparison of the two. We leave individual users of the data to decide which scale is most appropriate given their specific empirical questions and methodological preferences.

One of the longstanding criticisms directed at the PTS is that it does not disaggregate based on types of human rights violations.<sup>23</sup> When two (or more) countries receive the same score, this indicates that violations of physical integrity rights are roughly the same. However, the “mix” of human rights violations—the degree to which torture is carried out, the number of summary executions or extrajudicial killings, the size of a country’s political prisoner population, and so on—will invariably be different. As Poe et al. have explained, these various methods should be seen as substitutable policy options, and the application of one may prevent or render unnecessary the use of another.<sup>24</sup> To its credit, the CIRI has addressed this concern rather adeptly by disaggregating the constituent physical integrity violations considered by the PTS and coding them separately.

Disaggregating by abuse types has obvious benefits for researchers, and we especially welcome this valuable addition from the CIRI creators as it has allowed users to know the specific types of violence experienced by a country, which the PTS did not provide. However, what concerns us is the CIRI’s method of assembling these component parts into a complete picture of the human rights situation in the country. Specifically, we question the logic of summing these categories in order to establish this picture because in doing so users must make the assumption that an act of torture is equivalent to a disappearance, or an extrajudicial killing is equivalent to an instance of arbitrary imprisonment. We balk at this method because in our minds (and in line with logic of the PTS) it is possible to establish some rank order of the severity of the types of abuse. Our concern is particularly acute with regards to the CIRI creators’ assertion that the summed values that create the index show a progression from one

type of abuse to another. That is, the creators have asserted elsewhere that states proceed through abuse in a general sequence from political imprisonment to torture to killing to disappearance.<sup>25</sup> We remain unconvinced that any set sequence exists or that the summed CIRI *physint* score can accurately reveal this.

To illustrate these concerns we present an admittedly simple and perhaps exaggerated example of how summing abuses in the manner adopted by the CIRI differ from the scores generated by the PTS. Imagine that in country A, security officials storm a labor rally and kill 100 labor union members. In the country B, however, 100 labor union members are arrested, then tortured, and then killed. According to the approach of the PTS, the level of political violence in these two countries would essentially be the same because the level of abuse adopted by the regime is similar. The same number of persons was targeted for violence and the maximum level of violence inflicted on that population was equivalent. However, according to our understanding of the CIRI index, the human rights situation in the second state would be considerably worse than the first state (or, more accurately, its score would make it appear to be much worse). The reason is that each violation would be coded separately.<sup>26</sup> Thus, while the first state would have 100 “incidents” of extrajudicial killings, the second country would be experiencing 300 human rights violations: 100 cases of imprisonment + 100 cases of torture + 100 cases of extrajudicial killing. What should also be pointed out is that this same number would result if it involved 300 people, where 100 people were imprisoned, another 100 people were tortured and yet another group of 100 were simply killed. The broader point is that disaggregation might not only complicate matters, but it can potentially provide a misleading picture.

Another problem to which disaggregation lends itself is a pretense of precision and

accuracy that we are quite confident (based on years of coding) seldom exists. We use the prohibition against torture and cruel and unusual punishment as an example. CIRI provides three distinct categories for each type of violence—in this case torture—each based on the number of incidents that occurred. A score of 2 means that there has not been a single instance that year of either torture or cruel and unusual punishment; a 1 means that these violations have been carried out occasionally (between one and forty-nine times); and finally, a score of 0 indicates that torture and/or cruel and unusual punishment have been carried out frequently (fifty or more confirmed cases).

In our view, there are several limitations to this approach that have not been explicitly addressed by the CIRI creators. First, we are hard pressed to find *any* country where we could say with any degree of confidence that some certain number (X) were tortured. Since the USDS and AI reports seldom (if ever) make any mention of an exact number of incidents of torture, we question why a range of numbers is provided in the first place. The point is that measuring levels of human rights abuse is nowhere near as exact as this, and it borders on the misleading to pretend otherwise.

The second problem relates to the categories themselves. As noted earlier, the PTS is premised on the idea of providing a relative measurement of a country's human rights practices. Some states have excellent human rights records (level 1), some have horrible human rights records (level 5), but the vast majority of states will fall somewhere in between these two extremes (levels 2-4). By contrast, the CIRI categories do not provide the same relative perspective. Although there are three categories, the “best” score (category 2) is simply eliminated from consideration when there is a single instance of torture or cruel and unusual punishment. Note that even states with the most exemplary human rights records (e.g., New

Zealand, Denmark, and Canada) have been accused of engaging in such practices.

This leaves only two other categories. One of these (category 1) is for countries with one-forty-nine incidents of torture, and the other (category 0) is for countries with fifty or more incidents. The first thing that should be said is that there is no real theoretical justification for this number. Why would a country with fifty incidents of torture be placed in a different category than a country that has engaged in forty-nine incidents of torture? In addition, there is no indication that the CIRI index factors in the size of a country. In that way, fifty incidents of torture in China (current population: 1.3 billion) would be treated as being the same as fifty incidents of torture in a small country such as Sao Tome (current population: 150,000).

The larger issue, however, is that in our view this coding scheme does not necessarily allow for an adequate comparison of the extent of torture across countries. The problem is that a country where 50 people have been tortured receives the same score (0) as a country where 500, or 5000, or 50,000 (or more) have been subjected to this practice. The point is that this range is so wide and so great that it loses a great deal of meaning. In each one of these countries torture would be coded as being “frequent” under the CIRI index. But the real question is this: how frequent is “frequent”? In our view, there are enormous differences between “gross,” “systematic,” and “widespread” torture, on the one hand, and “reports” of torture or where the commission of torture is described as being “routine,” or “common,” or “regular.” Or to be more specific, there are vast differences in terms of human rights protection between countries where fifty-one people have been tortured and one where this number is in the hundreds or even thousands. We argue that the PTS makes a more serious effort to reflect these important differences.

A related concern is that in its attempt at precision and to focus on the reported number of

abuses, the CIRI score has neglected to address the range of the violence (e.g., who is targeted). Event counts of violence say nothing about who gets targeted. Range is an important dimension of physical integrity violation because it tells us something about the selectivity of the violence. While an event count reflects the frequency with which an abuse (or set of abuses) occurs, range tells us whether the government is indiscriminately torturing and killing a broad swath of its citizens or whether it is selectively targeting specific groups based on their actions and affiliations. While we would not argue that the latter is less repugnant than the former from a normative sense, we do see the validity in attempting to rank these two government strategies. Thus, all else equal, a state that selectively targets a single societal group for violence will generally receive a lower score than a state that broadly targets its victims. This is most apparent in the difference between levels 4 and 5 specified above. While category 4 countries may witness hundreds of killings or acts of torture, the abuse is typically directed at those persons who actively involve themselves in politics. By contrast, states in category 5 have expanded violence to the entire population, making no distinction between politically active and a political persons. Consequently, category 5 countries are not simply more violent than category 4 countries but are more arbitrary with respect to the targets of violence.

Something we also finding puzzling about the CIRI is that it at times seems to blur the line between the arbitrary precision of event counts and the more qualitative selectivity of abuse approach employed by the PTS. Specifically, the two may lead to quite different pictures of the abuse observed in the country. While the CIRI attempts to derive categorical codes for abuse based on the actual number of observed events, as stated above and acknowledged in the CIRI coding manual, these numbers often simply unavailable. In such cases, the manual instructs the coders to rely on language within the report, providing the guidance that certain adjectives—

*gross, widespread, systematic, epidemic, extensive, routine, regular*—should lead to a score of zero (torture is “*frequent*”).<sup>27</sup> In our view it is incorrect to make the assumption that the event counts necessarily convert clearly into qualitative categories. For example, we are somewhat puzzled at how exactly “50 or more” instance of abuse translates into “widespread” abuse. The point being, “widespread, systematic, and extensive” are qualitative terms that reflect to a certain degree the selectivity of abuse and/or its *relative* frequency, whereas counts of abuse are objective values that are constant across population size and do not reflect any inherent pattern of violence. Both approaches are valid (though as mentioned above we prefer the former), but we question the logic of mixing between the two. We are likewise concerned about the potential for creating inaccurate comparisons between those cases in which the score is based on a count and those in which the score is based on narrative descriptors.

Table 1 provides a few examples of the differences in coding that occur between the PTS and the CIRI. For each of the countries listed in first column of the table, we provide the PTS and CIRI scores as well as a sample of other states that were assigned the same score for that year by the respective datasets. For example, the PTS assigns the Philippines a score of "4" for the year 2000 while CIRI assigns a "0." In line with the logic of the PTS, other countries placed within the same category should exhibit similar characteristics. Accordingly, in 2000 the PTS assigned Israel, Cameroon, and Nepal the same score as the Philippines, meaning that these states all committed relatively similar acts of violence against the population. According to the CIRI's coding system, the Philippines is located in the same category as Iraq, Sri Lanka, and Afghanistan. In our view, while the abuses of the Philippines were severe, the range and intensity of the violations were not sufficiently severe to place it within the most abusive category. Moreover, we believe that abuses observed in the Philippines in that year were far

more similar to those in Israel compared to those in Afghanistan. Similarly, it is our opinion that the patterns of abuse observed in the United Kingdom in 1991 were more similar to those in Botswana than those in Uganda during the year.

**Table 1: Examples of Difference in PTS and CIRI Scoring**

| <b>Countries</b>   | <b>Year</b> | <b>PTS</b> | <b>CIRI</b> | <b>Others States in Same PTS Category</b> | <b>Other States in Same CIRI Category</b> |
|--------------------|-------------|------------|-------------|---|---|
| <b>Angola</b>      | 1986        | 5          | 3           | Afghanistan, South Africa, Suriname       | Poland, Senegal, South Korea              |
| <b>Philippines</b> | 2000        | 4          | 8           | Israel, Nepal, Cameroon                   | Iraq, Sri Lanka, Afghanistan              |
| <b>Morocco</b>     | 1990        | 3          | 7           | Albania, Mexico, Vietnam                  | Guatemala, Cambodia, Iran                 |
| <b>Chile</b>       | 1998        | 2          | 6           | Kuwait, Latvia, South Korea               | Kenya, China, Turkey                      |
| <b>U.K.</b>        | 1991        | 1          | 3           | New Zealand, Botswana, Denmark            | Uganda, Bangladesh, Vietnam               |

The table is intended to illustrate the differences between the PTS and the CIRI with respect to coding decisions and categorizations of state violence. However, it provides only a set of superficial and hand-selected examples of these differences. In the following section we examine these differences more systematically.

## V. Illustrating PTS-CIRI Differences

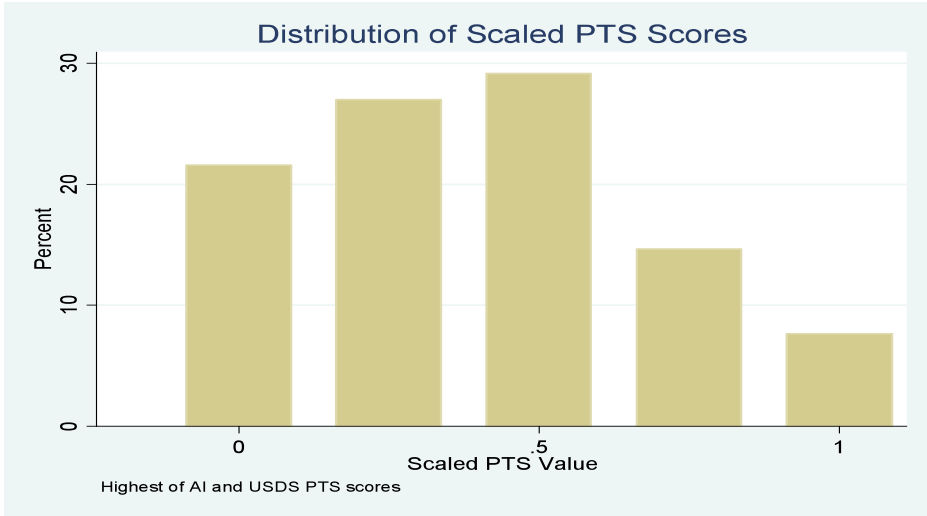
The above discussion provided a brief overview of the conceptual differences between the PTS and CIRI datasets. In this section we investigate these differences in a more systematic manner. We examine the extent to which the scores differ across the cases (n=3771) for which both scores are available. We then sample a few cases for which the scores differ significantly and

provide some explanations for these differences.

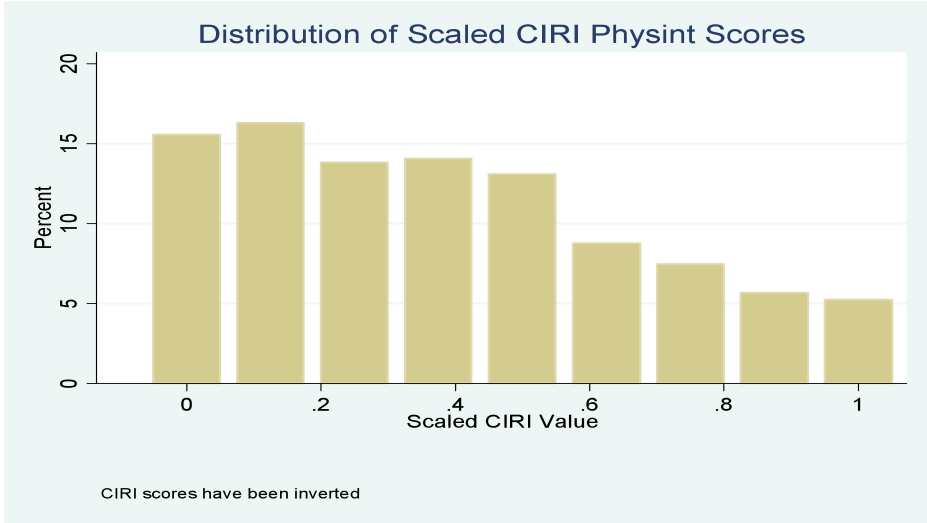
Because of the different categorical measures used by the two indices, we first scale the two measures.<sup>28</sup> In order to simplify comparisons on the two measures, we scale both to a 0-1 score. For the PTS the conversion equation is  $(PTS-1)/4$ . Consequently, we preserve the same five-category measure, but convert it to increments of 0.25. The CIRI score is first inverted to match the PTS, thus higher numbers reflect greater abuse. The inverted value is then scaled in the same way, only using increments of 0.125 to reflect its nine-point index. The CIRI conversion equation is  $[(physint * -1) + 8] / 8$ .

Figures 1 and 2 show the distributions of the scaled scores for each dataset. The similarities between the measures should be apparent. According to both measures, most countries rarely fall on the most abusive side of the scales. Over 48 percent of the PTS observations and nearly 46 percent of the CIRI observations score 0.25 or less on the scaled index. Both measures also suggest that, overall, states seldom fall into the most extreme categories of abuse. Yet despite some similarity in the distribution of the scores, some key differences emerge.<sup>29</sup> Relative to the CIRI, the PTS seems slightly less likely to place a country in the "best" or least abusive category. The PTS also tends more toward a normal distribution of state violence across the country years recorded here while the CIRI shows a slight bulge toward the less abusive (left) side of the scale. This is consistent with the intuition of the PTS mentioned above: some countries have great human rights performance; others have terrible performance; but most fall somewhere in between. Notably, in 23 percent and 18 percent of the PTS and CIRI observations, respectively, the scores meet or surpass the 0.75 category. Thus, the PTS seems slightly more critical overall.

## **Figure 1**



**Figure 2**

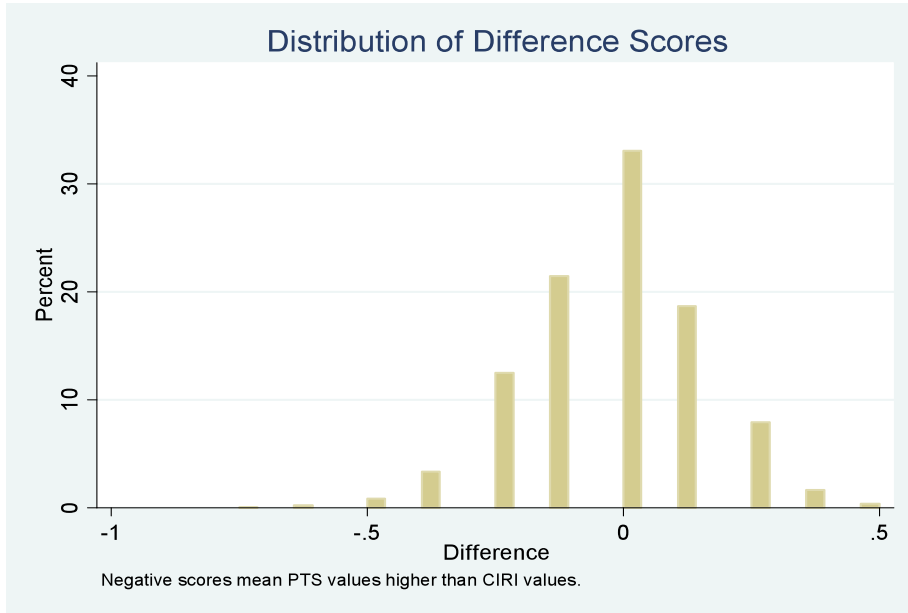


In order to assess the degree to which the scores differ for individual cases, we create a difference score by subtracting the PTS score from the CIRI physint score. Thus, negative values represent cases where the PTS rated a country as more abusive relative to the CIRI, and positive

scores reflect cases in which the CIRI is more critical. Figure 3 shows the distribution of the difference scores. According to this figure, the largest single difference category is 0, meaning that the two scores were identically matched along the 0-1 scale. This occurs in approximately 33 percent of cases. In addition, in nearly 73 percent of cases, the scaled scores were within 0.125 points of one another.

In our minds, 0.125 points would be tantamount to the same score. Many of these cases, we believe, would simply reflect the upper or lower bounds of the categories used in the PTS, those cases that are on the borders of the next higher or lower categories. The truncated PTS categories force coders to make hard decisions about the "location" of one state relative to another on a scale of violence. What should be understood about this construction is that while ideally each state within a category would exhibit exactly the same level of repressive violence against its citizens, in reality each category itself represents a range of behavior bounded by the descriptions of the categories presented in an earlier section. The CIRI allows for a finer aggregation, essentially doubling the number of total categories into which behavior can fall.

**Figure 3**

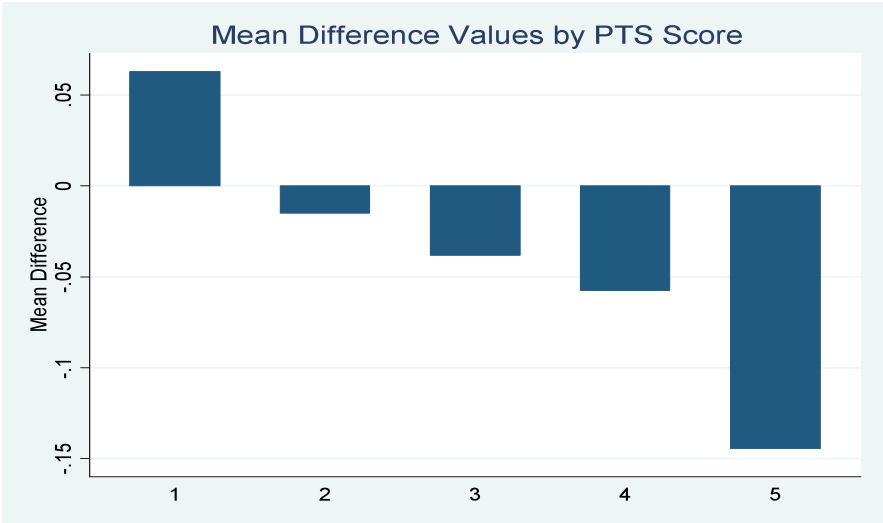


We believe as well that despite the assumed precision of the CIRI measure, the problem is perhaps worsened by the more strictly delimited component parts of the scale. For instance, each category of each component represents a clearly identified range of behavior. A state that scored a 1 in each category would be assumed to have killed, tortured, disappeared, and imprisoned between 1 and 49 persons. Yet, in our view, all else equal a state that committed each abuse a single time would be closer to a state that scored "2" (no abuses) in each category than a state that scores a "0", where there is no upper bound to the number of persons abused by the state. Despite increased precision, the CIRI scores do little to reflect this.

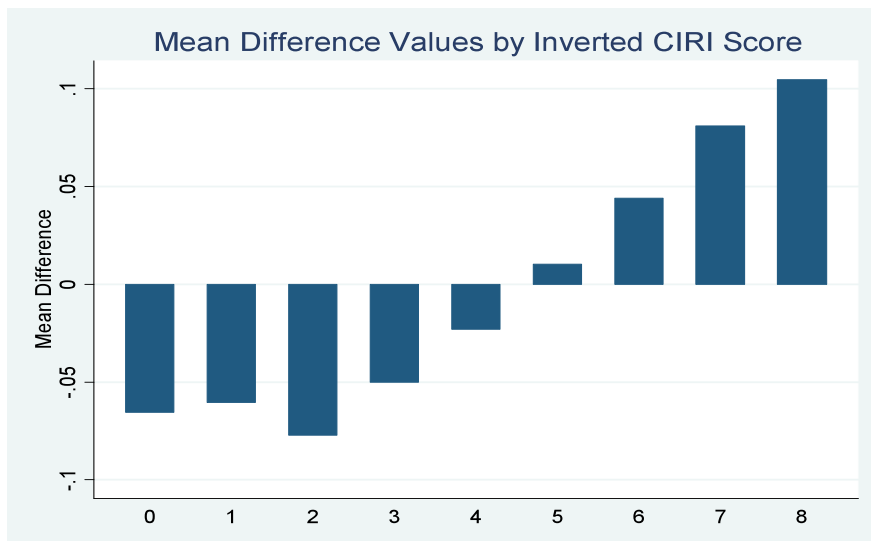
Some 27 percent of the scores show significant differences, differing by at least 0.25 points. This difference essentially represents one entire point on the PTS. A small number of cases (about 1.5 percent) have a difference score of at least 0.5 points or the equivalent of at least 2 points on the PTS. These are the cases we are interested in explaining. But where do the scores differ? It is possible that because of the different coding schemes and because CIRI disaggregates and sums scores, that differences cluster at either extreme of the scale. To assess

this, we compute the means of the difference scores by each category of the PTS and then the CIRI. Figures 4 and 5 show the distributions of the means. According to the data in the graphs, the absolute values of the means of the difference scores are larger at either extreme of both scales.<sup>30</sup> That is, the most extreme differences observed between the two scores occur either when a country is viewed as particularly abusive or particularly non-abusive by either of the scales. Difference values diminish significantly as the scores move toward the intermediate categories.

**Figure 4**



**Figure 5**



Overall, however, the mean difference scores are quite small across categories.<sup>31</sup> In almost every case the absolute value of the mean difference score is below or just exceeds 0.05. This means that the observed difference between the scores is only a fraction of one category of either score. Exceptions to this seem to be the most severely abusive cases as well as category "2" of the CIRI.<sup>32</sup> In Figure 4 (the graph examining means by PTS category) the mean difference score for category 5 is nearly -0.15, which we view as a substantial deviation from the CIRI scores. This is mirrored by Figure 5 (the graph reporting CIRI differences), which shows difference score means of nearly 0.1 and over 0.1 for the two most abusive CIRI categories. Consequently, it seems that the two scales differ most significantly in their coding (and understanding) of the most abusive states. The information displayed here suggests that on average when the PTS places a state in the most abusive category, the CIRI is placing it in a slightly less abusive category. The data suggest that on average when the PTS assigns a score of 5, the CIRI assigns a score of 7. What we find particularly interesting is that the same thing seems to occur for the CIRI. On average, the states it places in the most abusive category are ranked as less abusive by the PTS (less than one-half of a category). Consequently, the

differences in scores do not appear to result from one scale being harsher than the other. What is more likely is that the scales view abuse in distinct ways.

The notably high absolute value for category 2 of the inverted CIRI provides some speculative evidence of this. Interestingly, the mean value for category 2 is more in line with the difference scores for the most abusive categories. The negative mean difference score suggests that the PTS was generally more critical of the states in this category than was the CIRI. In attempting to account for this difference, it occurred to us that category 2 could represent a situation in which a state frequently engages in one of the abuses captured by the CIRI component categories but perpetrated none of the others. In this situation, a state's score might appear artificially low, yet the actual abuse could be particularly high. In examining the possible combination of CIRI scores, we found a few situations in which a state engages in significant amounts of one type of abuse but not others. Specifically, we noticed that in some cases, states engage in mass torture and yet not engage in any other abuses (n=95).<sup>33</sup> In these cases the mean difference scores match the mean difference for category 2 very closely (-0.82). Barring other violations, in cases of frequent or widespread torture the PTS would almost invariably code the country as more abusive than the CIRI. This sort of situation, we believe, may help explain where and why the scores differ.

Ideally we would develop a sophisticated statistical analysis to tease out these differences. However, given that both scales are premised on the same basic understanding of human rights abuse, and both are coded from the same source material, we feel that we would be unable to accomplish such an analysis without coding a diverse set of new variables based on the differences discussed in previous sections.<sup>34</sup> As an alternative, we sample a few of these cases and closely examine the original country reports from which the score are derived. From the

selected cases, we speculate that the most significant differences between the CIRI and the PTS country scores result from the following: different responsibility criteria, definitional differences, the (in our view) artificial constraints imposed by summing the disaggregated components of the CIRI, the PTS's attention to "range" as a conceptual element of abuse, and scaling for population size.

While both datasets attend to state violence, the PTS uses a broader definition of what constitutes violations of citizens' physical integrity by state agents. According to PTS coding guidelines, "state agents" extend to all those persons under nominal control of agencies of the state. Furthermore, the PTS extends "state" culpability to all actors with which the state (or its subsidiaries) has the capacity to exert significant influence. As discussed above, examples of this are militia or paramilitary organizations that work in association with or at the behest of the government, military, or some department therein. This is observed in a number of cases in which the PTS and CIRI differ by significant margins. For example, the 2005 scores for Guatemala differ by -0.5, suggesting that the PTS scored the country as significantly more violent than did the CIRI.<sup>35</sup> According to the component values of the CIRI, Guatemalan agents did not disappear any persons and held no political prisoners. However, it did commit between 1 and 49 acts of torture and extrajudicial executions.

According to the USDS report, "during the year the National Civilian Police (PNC) Office of Professional Responsibility (ORP) investigated twenty-four reports of police involvement in killings."<sup>36</sup> In addition, hundreds of killings were committed by non-state actors with ties to gangs, organized crime, private security firms, and "clandestine groups."<sup>37</sup> The report further states that police and security forces condoned or participated in some of these killings. While it is impossible to record the number of instances in which state personnel were involved,

let alone to what extent, the PTS considers this violence in its overall score. We have no indication that the CIRI attempts to account for this violence.

Definitional differences also influence deviations in the scores. Again using the case of Guatemala in 2005, the USDS report states that police personnel were involved in a number of kidnappings during the year. The CIRI score records no disappearances for the year. This is because the CIRI codebook informs coders not to count "typical" kidnappings as disappearances.<sup>38</sup> First we are unsure as to what constitutes a routine or typical kidnapping. More importantly, we believe that kidnappings, while perhaps not generally as grievous as politically motivated disappearances, are violations of an individual's physical integrity rights, especially given that victims often suffer physical abuse. The CIRI does not define what types of kidnapping might be included in their measure; nor do they instruct coders as to how to treat instances of abduction or disappearance in which state agents are complicit but not "directly" responsible. The PTS does include such violence because it relies on broader definitions of both responsibility and physical integrity violations.

Another difference between the scales is the PTS's attention to the range of the violence committed by the state. To illustrate how range affects the country scores between the two scales, consider the cases of Brazil and the Central African Republic (CAR). Table 2 provides comparison figures for these and other states. The difference scores for these countries in 2006 were 0 and -0.375 respectively. In 2006 the security forces of Brazil engaged in extrajudicial executions of hundreds of suspected drug dealers or other criminals. It received a 4 on the PTS. In the same year the military and other security forces of the government of the CAR killed hundreds of unarmed civilians during counterinsurgency operations. Because the CIRI does not consider the range of the violence—that is, who is targeted—it appears that Brazil is more

violent than the CAR.

While both countries committed similar numbers of extrajudicial killing in 2006, the Brazilian government is arguably more selective in its application of violence than the government of the CAR. According to the report for that year, the majority of victims of security force killings were suspected criminals (though it also targeted homeless persons, prostitutes, and street children). CAR security and military personnel, however, were less selective, killing scores of unarmed women and children without any discernible selection criteria other than their presence near a conflict zone. The PTS reflects this distinction while the CIRI does not.

**Table 2: Detailed Example of CIRI and PTS Differences**

| Countries         | Year | Difference | PTS | CIRI | Disap | Kill | Tort | Polpris |
|-------------------|------|------------|-----|------|-------|------|------|---------|
| <b>Brazil</b>     | 2006 | 0          | 4   | 6    | 1     | 0    | 0    | 1       |
| <b>CAF</b>        | 2006 | -0.5       | 5   | 5    | 2     | 0    | 0    | 1       |
| <b>East Timor</b> | 2006 | -0.5       | 4   | 2    | 2     | 1    | 2    | 1       |
| <b>Ethiopia</b>   | 1998 | -0.625     | 5   | 3    | 2     | 2    | 1    | 0       |
| <b>Guatemala</b>  | 2005 | -0.5       | 4   | 2    | 2     | 1    | 2    | 1       |
| <b>India</b>      | 2004 | 0.25       | 4   | 8    | 0     | 0    | 0    | 0       |
| <b>Tajikistan</b> | 1996 | 0.5        | 3   | 8    | 0     | 0    | 0    | 0       |
| <b>Yugoslavia</b> | 1994 | -0.5       | 5   | 4    | 2     | 1    | 1    | 0       |

An additional, closely related source of difference between the scores seems to be the constraints imposed by the CIRI's component parts. Hundreds of civilians were killed and scores were tortured or raped in the CAR that year; however, there were no reports of disappearances

and the state held few political prisoners.<sup>39</sup> Given the indiscriminate nature of the killings and tortures committed by state security forces, we believe that the number of disappearances or arbitrary detentions should do little to alter the score. Still, because the CAR did not disappear any persons and only held "numerous" political prisoners, which is interpreted by CIRI as the intermediate category, the state's cumulative score is only 3. Regardless of how many additional civilians over fifty were killed, and regardless of the dozens of rapes and tortures committed by government troops (likely well over fifty), the score cannot increase because it is constrained by the government's relative "restraint" in the areas of disappearances and prisoners.

A similar situation is observed in the difference between the scores for Yugoslavia (Serbia/Montenegro) in 1994. The CIRI assigned a score of 4 while the score from the PTS is a 5 (scale difference= -0.5). The difference can be explained in the extent of violence as understood by the PTS and the constraints imposed by CIRI categories. According to the AI report, thousands of persons were tortured or abused by Yugoslav authorities during the year, hundreds were detained, and numerous persons were killed (not to mention those killed in Serbian government-complicit violence in Bosnia).<sup>40</sup> The range of the violence is sufficient for the PTS to consider it among the most violent countries that year. However, because of the lack of disappearances and relatively few killings orchestrated by the state (in Serbia proper), the state received a mid-level abuse score from CIRI, the same score as Bulgaria, Morocco, and South Korea that year.

The difference between the scores for Ethiopia in 1998 (-0.625) provides a further example of this effect. In this year, the PTS assigned a 5 to Ethiopia while CIRI scored the country a comparatively less abusive 5. The CIRI score reflects that more than fifty persons were detained and that there were an intermediate number of tortures or similar abuses. By condensing

and packaging the scores in this way, the behavior of the Ethiopian government during its first year of conflict with Eritrea seems only moderately abusive. In fact, by this measure, Ethiopia's behavior was similar to that of Honduras, Jordan, and Bhutan in that year. However, a close reading of the source documents shows that the Ethiopian government continued to detain more than 10,000 political prisoners arrested during that and previous years. Moreover, the government forcibly expelled more than 40,000 men, women, and children under conditions equivalent to "ill treatment and abuse."<sup>41</sup> While the government disappeared or killed very few people in that year, the range of the abuse was quite significant, affecting a notable portion of the countries population.

This example raises an additional issue. The PTS considers ethnic cleansing and other mass detention and forcible expulsions a form of abuse or mistreatment. While not (exactly) the same as torture, we feel that violent expulsions fall under the category of physical abuse. As such, the PTS considers these events in its overall score. In the case of Ethiopia, this represents a mass violation of the rights of a large portion of the state's population. Not coding these violations ignores significant abuse conducted by the state.

The composite nature of the CIRI physint score also at times places states in a more violent category than does the PTS. For example, the difference score for Tajikistan in 1996 was 0.5. While the PTS assigned the country a score of 3, the same category as Albania, Mozambique, and Nicaragua, the CIRI scored the country a 0, placing it at the same level as Colombia, Sri Lanka, and the Sudan. The difference in the scores (and category peers) is that the CIRI's component categories sum to create a composite score. In this case, because the state (by the CIRI count) engaged in at least fifty of each type of abuse, it must be placed in the most abusive category. As we have pointed out earlier, the category itself says nothing about the

relative severity of the states within it. Thus, once a country passes the threshold number (50), it will receive the same score—no matter how great the violence (in this realm) becomes. In our view, Colombia was vastly more violent than Tajikistan that year, yet both states are given the same CIRI score. In fairness, we sympathize with this dilemma as the PTS has also been criticized for category truncation and the inability to differentiate between a really abusive state, say Colombia, and a genocidal state such as Rwanda in 1994. Still, we feel that despite its claim to greater precision, the CIRI is less capable of making these types of distinctions than is the PTS because it is constrained by precision, so to speak.

An additional explanation for the high difference scores relates to the PTS giving states the "benefit of the doubt" when the extent of violence is in dispute or unknown. For instance, the USDS report on Tajikistan states that there were "a significant number of extrajudicial killings."<sup>42</sup> In addition, it reports that civilians had been killed by artillery fire during the country's civil conflict. In neither case are fatality figures provided, nor does the report establish blame. We give the state the benefit of the doubt when ambiguous terminology such as this is used, especially when it is difficult to determine who was responsible. In addition, the report states that "a number of disappearances have been reported." However, it goes on to reveal that many of the persons believed to have been abducted were actually forcible state military conscripts. While we consider this type of abuse in generating the score, we do not believe that this behavior is the same as disappearances as they are traditionally defined. On the other two components of the scale, we concur with the CIRI scores. The state did widely abuse detainees and a large number of persons were arbitrarily detained for their political actions or beliefs. Our decision to code Tajikistan as a 3 is therefore based on our view that extralegal killing by the government was perhaps common but not extensive, there were few actual cases of politically

motivated disappearances, and there was frequent use of torture and arbitrary arrest. We view this as consistent with the description of category three presented above.

A final difference in the two scores involves the intensity of the violence *relative* to the population size. As discussed above, a raw count of abuse fails to capture the "risk" posed to the population or the actual pervasiveness of the violence observed in a country. In order to articulate this difference, the PTS subjectively scales abuse to the size of a country's population. Two examples of this stand out in our analysis of the data: East Timor (2006) and India (2004). In the first case, the country's small size suggests to us that we should be more sensitive to low levels of violence; in the latter, we feel that the enormity of the population should make us slightly less sensitive to small differences in the number of events.

The difference score for East Timor in 2006 was -0.5. The PTS score for that year was 4, while the CIRI assigned the relatively high (non-abusive) score of 6. The score for CIRI was the result of intermediate scores for killings and tortures in that year. According to the AI report, thirty-eight people were killed and 150,000 displaced during and in the aftermath of a police-military crackdown on a protest in the capital, Dili. The USDS report confirms much of this statement, though it reports that 30 persons were killed by "security forces and other actors."<sup>43</sup> While the number of deaths is relatively small to other countries that scored a 4 that year (Brazil, Nepal, the Philippines), we contend that the small population of the country makes the violence much more egregious because it affects a greater proportion of the country's citizens. To put this in perspective, the population of East Timor is roughly 900,000 persons. In our minds, thirty state killings in a country this size would be equivalent to hundreds of killings in a state with 9 million people (the population of Sweden or Bolivia for instance).

The reverse would be true of much larger countries. With a population of over a billion

people, security forces in India must commit greater numbers of abuses before they pose the same "threat" to the population as did the East Timorese forces—thirty some political killings would be proportionally tiny in India. The CIRI does not consider the number of abuses relative to the population. Thus, for most years India scores among the worst categories of the CIRI scale because state forces frequently commit at least fifty instances of each physical integrity category each year. While we agree that India is an abusive state, we disagree that it is so abusive to its population overall that it ranks beside states such as Afghanistan, Burma, and North Korea for most of the previous decade. The PTS score for India has sought to reflect that while abusive, Indian forces do exhibit constraint compared to many other countries. Consequently, it has received scores of either 3 or 4 on the PTS scale for all of the years for which a score is available.<sup>44</sup>

## VI. Conclusion

This research note has attempted to clarify some key elements of the PTS and to highlight a few differences between it and the more recent CIRI human rights scale. It has specifically attempted to provide users of the data with a more detailed account of what is being measured by the PTS and a more transparent look into the coding scheme it utilizes than what has previously been made available.<sup>45</sup> While this deeper look at the internals of the PTS may seem quite tardy to many users, we believe that something of a re-introduction and clarification is in order. We believe this is especially true given the advent of recent datasets such as the CIRI that offer similar measures of state respect for human rights but arrive at their conclusions through different arithmetic and with different conceptions of the severity of abuse.

We are indebted to the creators of such datasets for challenging us to review our own coding process and to closely examine where and why their measures differ from ours. Our comparisons, we believe, shed light on important differences between the two measures. Most quantitative human rights and political violence scholars have taken as given that the two scores capture state human rights behaviors in similar ways. We believe that overall this is true, but that scholars should be aware of the underlying causes of the differences that do exist. We do not advocate the PTS as the superior human rights indicator. Indeed, CIRI's disaggregation of abuse by type represents a significant advancement in human rights and repression data. We do believe, however, that the PTS better captures the relative severity of abuse across countries. We leave it to individual scholars to determine which measure best suits their needs.

## Endnotes

\* Reed M. Wood is a doctoral candidate in Political Science at the University of North Carolina—Chapel Hill. His research focuses on state repression, insurgency, and political violence. He has recently published articles in *International Studies Quarterly* and the *Journal of Peace Research*. His dissertation examines the causes of insurgent and state violence against non-combatants during civil conflicts.

\*\* Mark Gibney is the Belk Distinguished Professor at the University of North Carolina-Asheville. His most recent book publications include *International Human Rights Law: Returning to Universal Principles* (Rowman & Littlefield, 2008) and an edited volume (Gibney et al.) *The Age of Apology: Facing Up to the Past* (University of Pennsylvania Press, 2007). In addition, he has two forthcoming books. The first is the edited volume (Gibney & Skogly) *Universal Human Rights and Extraterritorial Obligations* (University of Pennsylvania Press) and the second is Sabine Carey, Mark Gibney and Steven Poe *The Politics of Human Rights: The Quest for Human Dignity* (Cambridge University Press).

---

<sup>1</sup> THE STATE AS TERRORIST: THE DYNAMICS OF GOVERNMENTAL VIOLENCE AND REPRESSION (Michael Stohl & George A. Lopez eds, 1984); GOVERNMENT VIOLENCE AND REPRESSION: AN AGENDA FOR RESEARCH (Michael Stohl & George A. Lopez eds, 1986); David Cingranelli & Thomas E. Pasquarello, *Human Rights Practices and the Distribution of US Foreign Aid to Latin American Countries*, 29 AM. J. POLITICAL SCIENCE 539 (1985); Raymond Duvall & Michael Stohl, *Governance by Terror*, in THE POLITICS OF TERRORISM 231-45 (Michael Stohl ed, 1983).

<sup>2</sup> LARS SCHOULTZ, HUMAN RIGHTS AND UNITED STATES POLICY TOWARD LATIN AMERICA (1981); David Carleton & Michael Stohl, *The Foreign Policy of Human Rights: Rhetoric and Reality from Jimmy Carter to Ronald Reagan*, 7 HUM. RTS. Q. 205 (1985); RAYMOND GASTIL, FREEDOM IN THE WORLD POLITICAL RIGHTS AND CIVIL LIBERTIES, 1980 (1980).

<sup>3</sup> Mark Gibney & Matthew Dalton, *The Political Terror Scale*, 4 POLICY STUD. & DEV. NATIONS 73 (1996); Mark Gibney, Reed Wood & Linda Cornett, *The Political Terror Scale* (2008), available at <http://www.politicalterrorsscale.org>. For an extensive bibliography see the PTS website *id.*: [www.politicalterrorsscale.org](http://www.politicalterrorsscale.org).

<sup>4</sup> David L. Cingranelli & David L. Richards, *Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights*, 43 INT'L STUD. Q. 401 (1999).

---

<sup>5</sup> Carleton & Stohl, *The Foreign Policy of Human Rights*, *supra* note 2; Michael Stohl, David Carleton, George Lopez & Stephen Samuels, *State Violations of Human Rights: Issues and Problems of Measurement*, 8 HUM. RTS. Q. 592 (1986).

<sup>6</sup> Steven Poe & Neal Tate, *Repression of the Human Right to Personal Integrity in 1980s: A Global Analysis*, 88 AM. POL. SCIENCE REV. 853 (1994); Steven C. Poe, C. Neal Tate & Linda Camp Keith, *Repression of the Human Right to Personal Integrity Revisited: A Global Cross-National Study Covering the Years 1976-1993*, 43 INT'L STUD. Q. 291 (1999).

<sup>7</sup> Carleton & Stohl, *The Foreign Policy of Human Rights*, *supra* note 2.

<sup>8</sup> See Duvall & Stohl, *supra* note 1; Stohl et al., *supra* note 5, at 594-95.

<sup>9</sup> (see AMNESTY INTERNATIONAL, AMNESTY INTERNATIONAL REPORT 2008: THE STATE OF THE WORLD'S HUMAN RIGHTS (2008), *available at* <http://report2008.amnesty.org>; UNITED STATES DEPARTMENT OF STATE, COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES (2008), *available at* <http://www.state.gov/g/drl/hr/>; Jorge Restrepo & Michael Spagat, *Civilian Casualties In the Colombian Conflict: A New Approach to Human Security*, (unpublished manuscript) Royal Holloway College, University of London (2004).

<sup>10</sup> We should note here that the year associated with the Amnesty International annual report is the publication year and not the year actually covered in the report. For example, the 2005 AI report would cover the events occurring in 2004. This has often led to confusion.

<sup>11</sup> See Steven C. Poe, Sabine Carey & Tanya Vazquez, *How are These Pictures Different? A Quantitative Comparison of the U.S. State Department and Amnesty International Human Rights Reports, 1976-1995*, 23 HUM. RTS. Q. 650 (2001) for a full analysis and discussion of biases in the reports and change in bias over time.

<sup>12</sup> In general, the PTS has been fortunate to have significant consistency among its coders. Mark Gibney has been the project manager and a principal coder since 1984. Consistency in coders is particularly important for coding schemes which rely on significant subjective assessment of qualitative information rather than event codes or other more objective criteria. This helps ensure consistency in the coding processes and minimizes the introduction of new biases that occur with coder changes.

<sup>13</sup> This might occur in one or two cases per year.

<sup>14</sup> GASTIL, *supra* note 2.

<sup>15</sup> Stohl et al., *State Violations of Human Rights*, *supra* note 5.

<sup>16</sup> It is difficult (and we believe largely inappropriate) to quantify scope. Doing so necessitates being able to count X number of imprisonments as equivalent to Y tortures and Z killings. However, it should be possible to rank order the severity of these abuses along an ordinal scale see Stohl et al., *id.* This is essentially the logic behind the PTS measure. The PTS does not

---

attempt to count and compare raw numbers of events.

<sup>17</sup> For instance, regime violence directed against labor leaders and political activists that results in hundreds of killings and disappearances would be more selective (have a more confined range) than indiscriminate violence toward apolitical peasants, even if the number of actual abuses in the latter case is lower. This is most obvious in the distinction between categories 4 and 5 (see discussion below), although it can certainly affect scores at any level of intensity.

<sup>18</sup> It might also be noted that abuses are rarely so effectively limited to the group or sub-group that is the primary target. In the Rwandan case, moderate Hutus as well as members of the Twa were extensively targeted for violence.

<sup>19</sup> David Cingranelli & David Richards, CIRI Human Rights Database (2009), *available at* <http://ciri.binghamton.edu>.

<sup>20</sup> The CIRI creators provided a more superficial account of some of these differences in an earlier publication. See Cingranelli, David L., & David L. Richards, *Measuring the Pattern, Level, and Sequence of Government Respect for Human Rights*, 43 INT'L STUD. Q. 407 (1999).

<sup>21</sup> The CIRI index also offers measures for a range of other human rights and civil liberties including union rights, women's rights, economic freedoms, and among others. See the dataset website for more information and a detailed codebook.

<sup>22</sup> This is Kendall's Tau-b coefficient. That the State Department PTS score is more highly correlated with the CIRI is largely due to the fact that CIRI uses the US Department of State reports as the primary source and the Amnesty International reports as a secondary source (David L. Cingranelli & David L. Richards, (CIRI) Human Rights Data Project Coding Manual, version 3.13.08, at 3 (2008). The Tau-b coefficient for the "worst" of the AI or USDS score is 0.74.

<sup>23</sup> James M. McCormick & Neil J. Mitchell, *Research Note: Human Rights Violations, Umbrella Concepts, and Empirical Analysis*, 49 WORLD POLITICS 510 (1997).

<sup>24</sup> Poe, et al., *Repression of the Human Right to Personal Integrity Revisited*, *supra* note 6, at 298.

<sup>25</sup> *Id.* at 21.

<sup>26</sup> See Cingranelli & Richards, CIRI Human Rights Data Project Coding Manual, *supra* note 22, at 5.

<sup>27</sup> *Id.* at 18.

<sup>28</sup> Unless otherwise noted, all comparisons use the "worst" PTS score from either the AI or USDS-derived scores.

<sup>29</sup> The distribution of scores for the period 1976-2006, the entire range for which the PTS is

---

available, is virtually identical. For scores based on the Amnesty International reports the distribution is quite similar; for the Department of State-derived scores there is a slight shift to the left. This superficially suggests that Amnesty is overall more critical of state performance than is the US State Department. For discussion see Poe et al., *How are These Pictures Different?*, *supra* note 11.

<sup>30</sup> CIRI scores have been inverted.

<sup>31</sup> We also calculated the mean difference values by year. The mean values show remarkable stability by year. In only 3 years does the absolute value of the mean difference exceed 0.05 (1983, 1986, and 2002). In these few cases, the mean values are each between -0.05 and -0.075.

<sup>32</sup> Throughout this article we refer to the inverted CIRI score we have constructed for purposes of comparison. As such, the inverted score of "2" to which we refer here reflects of score of "6" in the original CIRI coding system.

<sup>33</sup> The only other case was the less frequent scenario in which states engage in "frequent" political imprisonment but no other abuse (mean=-0.72).

<sup>34</sup> These would presumably be based on the differences cited above. As such, we would be required to essentially code for frequency, range, and other components that we feel are aspects of the PTS. We would also need to construct a database of militia and paramilitary groups. We believe this would represent an immense time commitment, and one that is not necessarily warranted for such generally similar scales.

<sup>35</sup> The CIRI score places Guatemala in the same category as Singapore, Canada, Greece, and Spain in that year. The PTS, by contrast, places it in the same category as India, Chad, Egypt, and Brazil.

<sup>36</sup> UNITED STATES DEPARTMENT OF STATE, COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES, "GUATEMALA" (2006)

<sup>37</sup> *Id.* at 36.

<sup>38</sup> CIRI codebook informs coders not to count "typical" kidnappings as disappearances. Cingranelli & Richards, CIRI Human Rights Data Project Coding Manual, *supra* note 22, at 14).

<sup>39</sup> UNITED STATES DEPARTMENT OF STATE, COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES, "CENTRAL AFRICAN REPUBLIC" (2008)

<sup>40</sup> AMNESTY INTERNATIONAL, AMNESTY INTERNATIONAL REPORT 1995: THE STATE OF THE WORLD'S HUMAN RIGHTS, "YUGOSLAVIA" (1995)

<sup>41</sup> AMNESTY INTERNATIONAL, AMNESTY INTERNATIONAL REPORT 1999: THE STATE OF THE WORLD'S HUMAN RIGHTS, "ETHIOPIA" (1999); UNITED STATES DEPARTMENT OF STATE, COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES, "ETHIOPIA" (1999).

---

<sup>42</sup> UNITED STATES DEPARTMENT OF STATE, COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES, “TAJKISTAN” (1997)

<sup>43</sup> Both reports also suggest abuse or excessive force were used by state security personnel. This information figured in the PTS scoring for the country, but it is notably absent from the CIRI component scores.

<sup>44</sup> AI scores have more often trended toward 3 while USDS scores trend toward 4.

<sup>45</sup> See Gibney & Dalton, *The Political Terror Scale*, *supra* note 3, for an earlier discussion.