

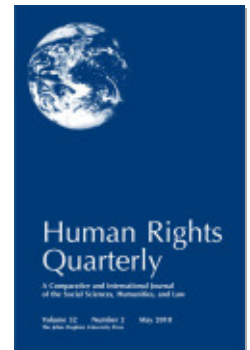


PROJECT MUSE®

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

Reed M. Wood, Mark Gibney

Human Rights Quarterly, Volume 32, Number 2, May 2010, pp. 367-400 (Article)



Published by Johns Hopkins University Press

DOI: <https://doi.org/10.1353/hrq.0.0152>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/380491>

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

* 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. He is currently co-manager for the Political Terror Scale and has coded for the PTS since 2000.

** 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 two edited volumes *The Age of Apology: Facing Up to the Past* (Mark Gibney et al. eds., Univ. of Pennsylvania Press 2007) and *Universal Human Rights and Extraterritorial Obligations* (Mark Gibney & Sigrun Skogly eds., Univ. of Pennsylvania Press 2010). In addition, he has a forthcoming book: Sabine Carey, Mark Gibney & Steven Poe, *The Politics of Human Rights: The Quest for Human Dignity* (Cambridge Univ. Press 2010). Gibney has managed the coding for the PTS since 1984.

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.¹ 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.² 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.³

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).⁴ 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

-
1. 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. POL. SCI. 539 (1985); Raymond D. Duvall & Michael Stohl, *Governance by Terror*, in THE POLITICS OF TERRORISM 179 (Michael Stohl ed., 1983).
 2. LARS SCHULTZ, 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 D. GASTIL, FREEDOM IN THE WORLD: POLITICAL RIGHTS AND CIVIL LIBERTIES (1980).
 3. Mark Gibney & Matthew Dalton, *The Political Terror Scale*, 4 POL’Y STUD. & DEV. NATIONS 73 (1996). For an extensive bibliography see *The Political Terror Scale*, available at <http://www.politicalterror scale.org>.
 4. For a description of the CIRI project, see David L. Cingranelli & David L. Richards, *Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights*, 43 INT’L STUD. Q. 407 (1999); David Cingranelli & David Richards, CIRI Human Rights Data Project, available at <http://ciri.binghamton.edu/index.asp>.

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. PTS: WHAT IT MEANS AND WHAT IT MEASURES

The PTS is a standards-based human rights data set first created by a group of colleagues at Purdue University in the early 1980s.⁵ 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 that they have done using data from the Political Terror Scale.⁶ However, there has never been a Purdue Terror Scale or a Poe-Tate Scale—there has only ever been the Political Terror Scale.

The original version of the PTS coded fifty-nine countries for the years 1976–1983.⁷ 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 states' human rights practices during more than three decades (1976–2008).

Another 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. The meaning of the word terror that we employ is, of course, different from the meaning that “terror” has now taken on, particularly after the events of 11 September 2001.

-
5. Carleton & Stohl, *supra* note 2, 211–18 (using data to compare the Carter Administration's respect for human rights to the Reagan Administration's respect for human rights); Michael Stohl, David Carleton, George Lopez & Stephen Samuels, *State Violation of Human Rights: Issues and Problems of Measurement*, 8 HUM. RTS. Q. 592 (1986) [hereinafter *State Violation of Human Rights*].
 6. For examples of Steven Poe and Neal Tate's work, see Steven C. Poe & C. Neal Tate, *Repression of the Human Right to Personal Integrity in 1980s: A Global Analysis*, 88 AM. POL. SCI. 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).
 7. Carleton & Stohl, *supra* note 2.

In addition, the PTS measures actual violations of physical integrity rights more than it measures general political repression. In fact, there will be instances where a government is so repressive that, as a consequence, there are relatively few acts of political violence. For a state that is truly efficient in its use of coercive repression against its citizens, the repression from an earlier period will continue to repress citizens or deter challenges to the controlling regime in subsequent periods.⁸ For example, the former Soviet Union received a PTS score of either 2 or 3 in each year in the early 1980s—scores that certainly do not reflect the overall level of political repression and social control employed by the totalitarian regime during that period.⁹ Yet, the fact that the USSR had engaged in massive organized violence against its population during earlier periods, coupled with the state's ability to monitor and police its population, meant that the USSR did not need to resort to high levels of explicit violence during that time in order to keep its population repressed.

The PTS focuses on state behavior. As such, domestic (family) or societal (mob, clan) violence, which 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, our rationale is that private actors carry out this type of violence, 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). Moreover, 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 determining the state's score; yet, security forces committing extrajudicial execution and abusing individuals in custody while prosecuting these criminal drug syndicates will be considered in Mexico's score.

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 for the PTS over the years has been 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

8. See Duvall & Stohl, *supra* note 1; *State Violation of Human Rights*, *supra* note 5, at 594–95.

9. The PTS scores for the Soviet Union were as follows: 1980 (SD 3, AI 3); 1981 (SD 3, AI 3); 1982 (SD 3, AI 3); 1983 (SD 3, AI 3); 1984 (SD 2, AI 2); 1985 (SD 2, AI 2); 1986 (SD 3, AI 3). The Political Terror Scale, available at <http://www.politicalterrorscale.org>.

killing of civilians by security forces during combat, and other arbitrary deprivation of life by state actors. The PTS does not include state-sanctioned executions that occur after trials that conform to international standards. 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 killings that take place after legal proceedings.

On some occasions, it is not clear if the state is the actor directly responsible for a given abuse. This is particularly a problem when paramilitary organizations or local militias engage in a significant amount of violence. The extent of government involvement in the activities of paramilitary groups is not easily identifiable—governments often choose to allow these organizations to operate for this reason. It is therefore incumbent upon coders to make these decisions in terms of how to properly code such situations. The reports from which the scores are generated often provide some insight into the level of involvement; thus, coders use these reports to identify a state actor level of involvement in the human rights abuses in a particular country.

How coders have dealt with Colombia is a perfect example of this process. Gross human rights violations by official state security forces have diminished markedly since the early 2000s in Colombia.¹⁰ At the same time, violence committed by paramilitary organizations has not shown 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.¹¹ 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 Colombian government and military provide support to, or are complicit in, the violence attributed to armed groups such as the AUC.¹² As a result, the PTS scores provided for Colombia are

-
10. Quite typical from early in the decade is this language from the 2002 USDS Report on Colombia: “A small percentage of total human rights abuses reported were attributed to state security forces; however, some members of the government security forces continued to commit serious abuses, including unlawful and extrajudicial killings. Some members of the security forces collaborated with paramilitary groups that committed serious abuses. Impunity remained at the core of the country’s human rights problems.” U.S. Dep’t of State, 2002 Country Reports on Human Rights Practices, Colombia (2003), available at <http://www.state.gov/g/drl/rls/hrrpt/2002/18325.htm>.
 11. See AMNESTY INT’L, AMNESTY INTERNATIONAL REPORT 2008: THE STATE OF THE WORLD’S HUMAN RIGHTS 97 (2008), available at <http://archive.amnesty.org/report2008/document/101.pdf>; U. S. Dep’t of State, 2008 Country Reports on Human Rights Practices: Colombia (2009), available at <http://www.state.gov/g/drl/rls/hrrpt/2008/wha/119153.htm>; Jorge Restrepo & Michael Spagat, Civilian Casualties in the Colombian Conflict: A New Approach to Human Security (27 Oct. 2004) (unpublished manuscript, on file with Royal Holloway College, University of London), available at http://eprints.rhul.ac.uk/439/1/HS_in_Colombia_Civil_Conflict.pdf.
 12. U.S. Dep’t of State, 2008 Country Reports on Human Rights Practices, *supra* note 11; AMNESTY INT’L, *supra* note 11.

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;” in which security personnel have provided material support for the abduction of political figures or journalists; or in which security personnel have been complicit in kidnappings for profit. In sum, notwithstanding some unfortunate confusion regarding its name (confusion that we hope will now be laid to rest), the PTS provides more than three decades of data on the human rights practices of states. Moreover, as we will describe in the following section, this data has been both consistent and reliable.

III. CODING THE PTS

The data for the coding comes from two annual sources: the US Department of State *Country Reports on Human Rights Practices* and the Amnesty International *Annual Report*.¹³ In the construction of the index for each year, PTS coders are instructed to presume that the information in the reports is accurate and complete. Thus, any biases found in the annual reports of the two organizations should be evident in the PTS indices.¹⁴ Each country in each report is coded by at least two senior coders,¹⁵ 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, and when there is a discrepancy, the difference has almost always been a single level. 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 consulted. Bringing in a third coder seldom occurs in practice.¹⁶

13. 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 Amnesty report would cover the events occurring in 2004.

14. See Steven C. Poe, Sabine C. Carey & Tanya C. Vazquez, *How Are These Pictures Different? A Quantitative Comparison of the US State Department and Amnesty International Human Rights Reports, 1976–1995*, 23 HUM. RTS. Q. 650 (2001) [hereinafter *How Are These Pictures Different?*] for a full analysis and discussion of biases in the reports and change in bias over time.

15. In general, the PTS has been fortunate to have had 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 that 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. Political Terror Scale, Frequently Asked Questions, available at <http://www.politicalterrorscale.org/faq.php>.

16. The need to consult a third coder might occur in one or two cases per year.

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.¹⁷

While the descriptions of the coding categories are well known to most PTS users, perhaps few scholars are 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.¹⁸ In brief, *scope* refers to the type of violence being carried out by the state (imprisonment, torture, killing, etc.).¹⁹ *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.²⁰

17. *How Are These Pictures Different?*, *supra* note 14 at 658.

18. *State Violation of Human Rights*, *supra* note 5, at 600–03.

19. *See id.* It is difficult (and 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, ranking the severity of these abuses along an ordinal scale is possible. This ranking is essentially the logic behind the PTS measure; the PTS does not attempt to count and compare raw numbers of events.

20. 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 on the PTS, although it can certainly affect scores at any level of intensity.

As the above discussion suggests, the PTS relies heavily on 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 occur 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 must 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, she should code that country a 2.

Perhaps the most elementary rule of all is that the coders need to understand that the PTS represents a continuum of human rights practices. In the end, 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 similarities to other countries with the same score. All of the states within a given category (score) should be characterized by similar patterns of abuses. Clearly, common sense is enormously important to this enterprise, and the PTS tries to avoid a mechanical application of the coding scale that arrives at absurd results. For example, a level 5 score might mean that the political violence in a country is extended to the entire population of the country. However, a situation in which gross and systematic human rights violations are being directed at a specific group—Rwandan Tutsis in 1994 or ethnic Albanians in Kosovo in 1999—should not result in a lower (better) score because not all of a country's population are equally at risk.

While this might seem to contradict the above discussion of range, two points pertaining to this situation are worth noting. First, although the target population may be delimited along ethnic or religious lines, the state's use of violence against such subsections of the society falls particularly high on the two other dimensions (scope and intensity) that underlie the logic of the PTS. In the above examples, while an ethnic minority (and not the whole population) was the immediate target of the state-sanctioned violence, the government employed violence with such frequency and brutality that both countries received scores of 5.²¹ Second, and perhaps more important, range is intended to differentiate among abused groups based on the groups' ob-

21. These cases are extraordinary—abuses are rarely so effectively limited to the group or sub-group that is the primary target.

servable actions, behaviors, or associations. That is, the PTS would assign a lower score to a state that was responsible for killing hundreds of political activists, labor leaders, and protesting students compared to a state that was responsible for executing the same number of apolitical peasants. This 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 Human Rights Project.²² 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.²³ In this section, we seek to address these differences in some detail.

Like the PTS, the CIRI measures state-sponsored violations of the subset of human rights known as physical integrity rights.²⁴ In addition, both projects derive their respective categorical scores by subjectively coding the same source data. Unsurprisingly, the two measures are rather highly correlated—approximately 0.73 for the USDS-derived PTS scores and 0.65 for AI scores.²⁵ That both the PTS and CIRI focus on the same types of violence, code from the same descriptive data, and come to many of the same results suggests a high degree of similarity between the datasets.

Despite the similarities, the two datasets exhibit notable differences. First, CIRI explicitly disaggregates physical integrity violations into several

22. David Cingranelli & David Richards, CIRI Human Rights Documentation, *available at* <http://ciri.binghamton.edu/documentation.asp>.

23. The CIRI creators provided a more superficial account of some of these differences in an earlier publication. See David L. Cingranelli & David L. Richards, *Measuring the Level, Pattern, and Sequence of Government Respect for Human Rights*, *supra* note 4, at 407–08, 414–15.

24. DAVID L. CINGRANELLI & DAVID L. RICHARDS, THE CINGRANELLI-RICHARDS CIRI HUMAN RIGHTS DATA PROJECT CODING MANUAL, version 7.30.08 (2008) [hereinafter PROJECT CODING MANUAL], *available at* http://ciri.binghamton.edu/documentation/ciri_coding_guide.pdf. The CIRI index also offers measures for a range of other human rights and civil liberties including union rights, women's rights, and economic freedoms, among others. DAVID L. CINGRANELLI & DAVID L. RICHARDS, SHORT VARIABLE DESCRIPTIONS FOR INDICATORS IN THE [SIC] THE CINGRANELLI-RICHARDS (CIRI) HUMAN RIGHTS DATA SET, version 12.07.08 (2008) [hereinafter SHORT VARIABLE DESCRIPTIONS], *available at* http://ciri.binghamton.edu/documentation/ciri_variables_short_descriptions.pdf.

25. 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. SHORT VARIABLE DESCRIPTIONS, *supra* note 24, at 3. The Tau-b coefficient for the "worst" of the Amnesty or USDS score is 0.74.

(though not necessarily all) of its component parts: disappearances, killing, torture, and imprisonment.²⁶ While PTS provides one score into which multiple dimensions of abuse have been collapsed, CIRI offers a 9-point scale constructed by summing scores from four subcomponents capturing the specific abuses mentioned above. Unlike the PTS, in which higher scores reflect *worse* state human rights practices, higher CIRI scores reflect *better* respect for these rights. Thus, while a “5” represents a worst-case scenario on the PTS, a “0” represents the worst-case scenario on the CIRI. However, for the remainder of this article, we invert the CIRI scores for ease of comparison. Henceforth, “8” reflects the most abusive score on the combined CIRI Physical Integrity score (physint) and “2” represents the worst score for each of the four subcomponents.

Second, CIRI 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. This section discusses these differences and then offers 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.²⁷ When two (or more) countries receive the same score, this indicates that violations of physical integrity rights are roughly the same in those countries. 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, etc.—will invariably be different. As Poe et al. explain, these various methods should be seen as “substitutable policy options, and that the choice of one may prevent or render unnecessary the use of the other.”²⁸ 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

26. SHORT VARIABLE DESCRIPTIONS, *supra* note 24, at 3–4.

27. James M. McCormick & Neil J. Mitchell, *Human Rights Violations, Umbrella Concepts, and Empirical Analysis*, 49 *WORLD POL.* 510, 511–15 (1997).

28. Poe, Tate & Keith, *Repression of the Human Right to Personal Integrity Revisited*, *supra* note 6, at 298.

as it has allowed users to know the specific types of violence experienced within a particular country, which the PTS does 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 to establish this picture because, in doing so, users must make the assumption that an act of torture is equivalent to a disappearance or that an extrajudicial killing is equivalent to an instance of arbitrary imprisonment. Our concern is particularly acute with regard 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 that states proceed through abuse in a general sequence from political imprisonment to torture to killing to disappearance.²⁹ We remain unconvinced that any such sequence exists or that the summed CIRI score (physint)³⁰ 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 Country B, however, 100 labor union members are arrested and imprisoned, 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. 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 (at least its score would make it appear to be much worse) because each violation would be coded separately.³¹ Thus, while the first state would have 100 incidents of extrajudicial killings, the second country would be responsible for 300 human rights violations: 100 cases of imprisonment + 100 cases of torture + 100 cases of extrajudicial killing. Moreover, this same number would result if Country B's situation 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)

29. David L. Cingranelli & David L. Richards, *Measuring the Level, Pattern, and Sequence of Government Respect for Human Rights*, *supra* note 4, at 411–15.

30. Physint stands for Physical Integrity Rights Index. It is the total score that the CIRI gives to a country after adding together the scores from four indicators: torture, extrajudicial killing, political imprisonment, and disappearance. *SHORT VARIABLE DESCRIPTIONS*, *supra* note 24, at 3.

31. See *PROJECT CODING MANUAL*, *supra* note 24, at 5.

seldom exists. We use the prohibition against torture and cruel and unusual punishment as an example for this assertion. CIRI provides three distinct categories for each type of violence—in this case, torture—each category based on the number of incidents that occurred.³² A score of 0 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 2 indicates that torture or cruel and unusual punishment have been practiced frequently (fifty or more confirmed cases).³³

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 a state tortured a certain number of individuals. Because 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. Measuring levels of human rights abuse is not an exact science, and pretending otherwise borders on the misleading.

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) whereas 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 (here, “0”) 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 through forty-nine incidents of torture, and the other (category 2) is for countries with fifty or more incidents. First, CIRI provides 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 the same as fifty incidents of torture in a small country, such as Sao Tome (current population: 150,000).

32. *Id.* at 18 (dividing the torture indicator into three categories with three corresponding scores).

33. Again to avoid confusion, throughout this paper we use an inversion of the CIRI scale. Consequently, CIRI's worst category is a “0”, suggesting that no physical integrity rights were respected. Herein, we convert this score to an “8” for ease of comparison with the PTS.

The larger issue, however, is that this coding scheme does not necessarily allow for an adequate comparison of the extent of torture across countries. A country that tortures fifty people receives the same score as a country that subjects 500, 5,000, or even 50,000 individuals to torture. The CIRI range is so wide and so great that it loses a great deal of meaning. For each one of these countries, the CIRI index would code torture as "frequent." But what does "frequent" really mean? In our view, there are enormous differences between "gross," "systematic," and "widespread" torture, on the one hand, and "reports" of torture or "routine," "common," or "regular" torture, on the other. In other words, a country where fifty people have been tortured better protects human rights than a country that has tortured hundreds, even thousands, of human beings—and countries' human rights scores should reflect this. We argue that the PTS makes a more serious effort to reflect these important differences.

CIRI's attempt at precision and to focus on the reported number of abuses is a related concern, for the CIRI score neglects to address the range of the violence. Counts of violence say nothing about who gets targeted. Range is an important dimension of physical integrity violations because range illustrates 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 do not argue that the latter is less repugnant than the former in a normative sense, we do find validity in attempting to rank these two government strategies. A state that selectively targets a single societal group will generally receive a lower (better) score than a state that broadly targets its victims. This is most apparent in the difference between levels 4 and 5. While a category 4 country's population 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 their perpetration of violence to the entire population, making no distinction between politically active and apolitical persons. Consequently, category 5 countries are not more violent than category 4 countries but are more arbitrary with respect to the targets of violence.

The CIRI also seems puzzling because it, at times, seems to blur the line between the arbitrary precision of event counts and the more qualitative selectivity employed by the PTS. Specifically, the two methods 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 acknowledged in the CIRI coding manual, these numbers are often simply unavailable. In such cases, the CIRI manual instructs its coders to rely on language within the reports, explaining that

certain adjectives—*gross, widespread, systematic, epidemic, extensive, routine, regular*—should lead to a score of 0 (torture is “practiced frequently”).³⁴ It is incorrect to make the assumption that the event counts necessarily convert into qualitative categories. For example, we are somewhat puzzled by how “50 or more” instances of abuse translate into “widespread” abuse. “Widespread, systematic, and extensive” are qualitative terms that reflect, to a certain degree, the selectivity of abuse and 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 we prefer the former), but we question the logic of mixing 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 the 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 has placed the country in the “worst” category 8. 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 approximately the same levels of political violence. 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.

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 dif-

34. See PROJECT CODING MANUAL, *supra* note 24, at 18.

Table 1
Examples of Difference in PTS and CIRI Scoring

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

Note: CIRI scores have been inverted for comparison. Herein, an “8” in the “CIRI” category represents the worst possible score while a “0” reflects the absence of each of the types of abuse considered.

ferences 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.³⁵ In order to simplify the comparisons of the two measures, we scale both datasets 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 illustrate 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, key differences emerge.³⁶ 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.

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 summed CIRI 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

35. Unless otherwise noted, all comparisons use the “worst” PTS score from either the AI or USDS-derived scores.

36. 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 AI is overall more critical of state performance than is the US State Department. For discussion see *How Are These Pictures Different?*, *supra* note 14.

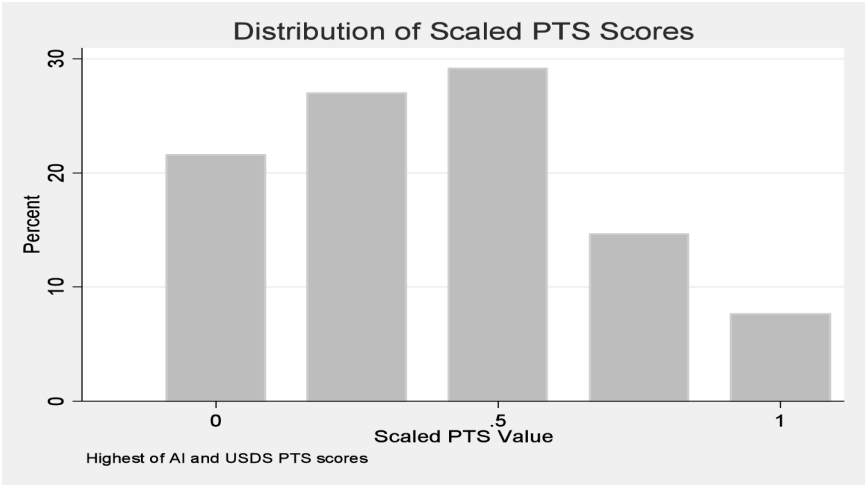


Figure 1

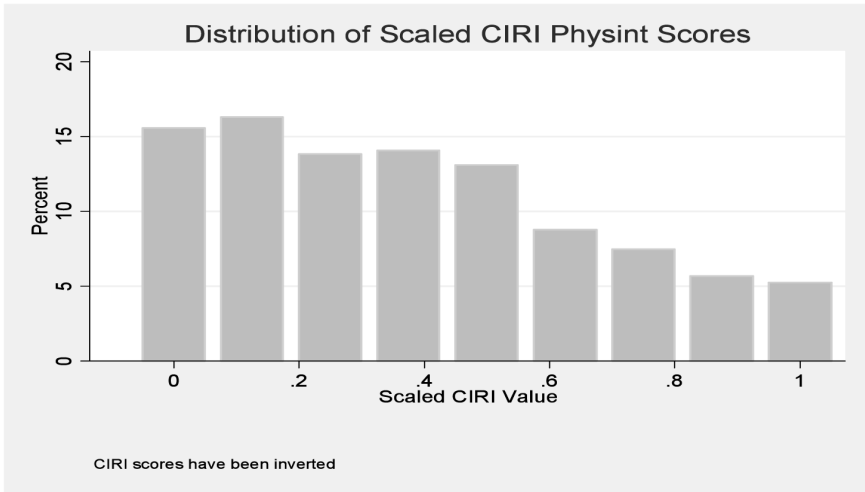


Figure 2

the difference scores. According to this figure, the largest single difference category is 0, meaning that the PTS and CIRI 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, a difference of 0.125 points is tantamount to the same score. Many of these are cases that are on the borders of the next higher or lower categories on the PTS scale. The truncated PTS categories force coders to make hard decisions about the “location” of one state relative to another on a scale of violence. 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 bound by the descriptions of the categories presented in Section III. The CIRI allows for a finer aggregation, essentially doubling the number of total categories into which behavior can fall.

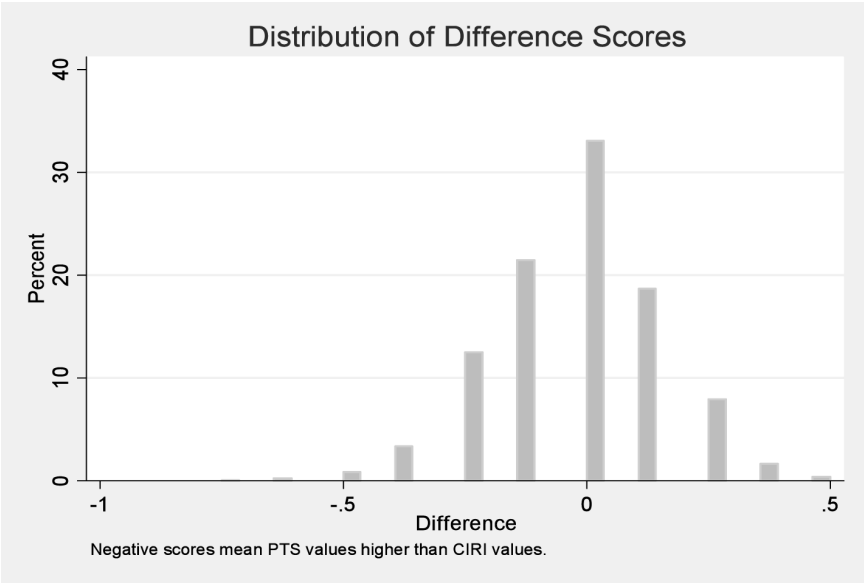


Figure 3

However, what also has to be recognized is that there will be cases where the assumed precision of the CIRI index will lead to odd or skewed results. For instance, each category of each component represents a clearly

identified range of behavior. One could assume that a state that scored a “1” in each category would have killed, tortured, disappeared, and imprisoned between 1 and 49 persons. In our opinion, a state that committed each abuse a single time would be closer to a state that scored “0” (no abuses) in each category than a state that scores a “2,” where there is no upper bound to the number of persons abused by the state.

Some 27 percent of the scores show significant differences, varying 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, the equivalent of at least 2 points on the PTS. 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. 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.

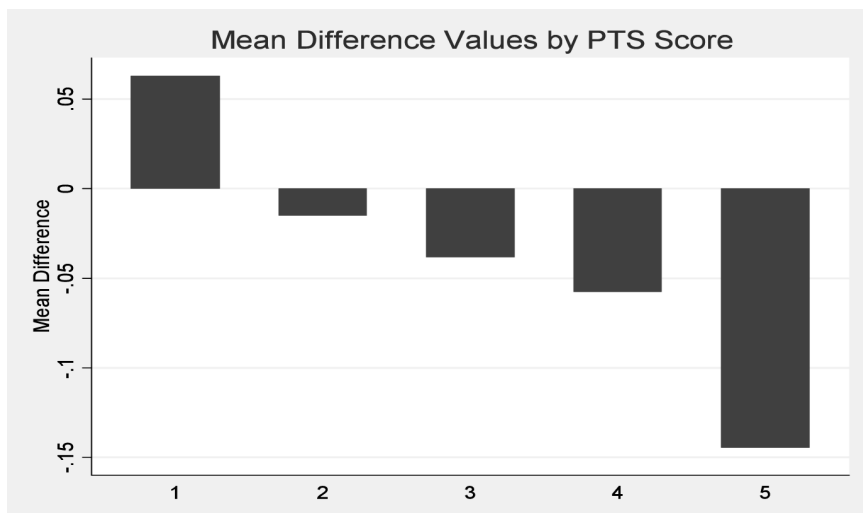


Figure 4

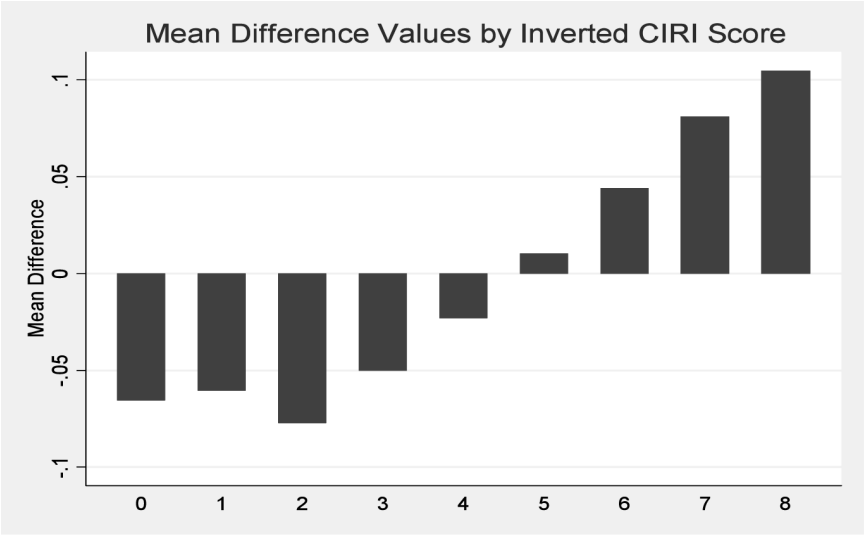


Figure 5

Overall, however, the mean difference scores are quite small across categories.³⁷ 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. 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, 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 also suggest that on average, when the PTS assigns a score of 5, the CIRI assigns a score of 7 (just shy of the worst possible score). 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 slightly less abusive by the PTS (though by less than

37. We also calculated the mean difference values by year. The mean values show remarkable stability by year. In only three 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.

one-half of a category). Consequently, the differences in scores do not appear to result from one scale being harsher than the other. More likely, the scales view abuse in distinct ways.

The notably high absolute value for category 2 of the inverted CIRI scale provides some speculative evidence of these distinctions. 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 perpetrates none of the others. In this situation, a state's score might appear artificially low, whereas 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 but do not perpetrate other abuses ($n=95$).³⁸ 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 will almost invariably code the country as more abusive than the CIRI. This sort of situation helps to 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.³⁹ As an alternative, we sample a few of these cases and closely examine the original country reports from which the scores 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 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

38. The only other case was the less frequent scenario in which states "frequently practiced" political imprisonment but no other abuse (mean= -0.72).

39. 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. This would represent an immense time commitment, and one that is not necessarily warranted for such generally similar scales.

agents. Under the PTS approach, the definition of state agents extends to all those persons under nominal control of agencies of the state. Thus, the PTS extends state culpability to all actors on which the state (or its subsidiaries) has the capacity to exert significant influence. Examples of this type of state actor include militia or paramilitary organizations that work in association with or at the behest of the government, military, or some department therein. These kinds of groups influence 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.⁴⁰ 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 2005 USDS report on Guatemala, "[d]uring the year the National Civilian Police (PNC) Office of Professional Responsibility (ORP) investigated 24 reports of police involvement in killings."⁴¹ In addition, hundreds of killings were committed by non-state actors with ties to gangs, organized crime, private security firms, and "clandestine groups."⁴² 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.⁴³ The CIRI score records no disappearances for the year because the CIRI codebook informs coders not to count "typical" kidnappings as disappearances.⁴⁴ First, we are unsure as to what constitutes a 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 because victims often suffer physical abuse. The CIRI does not define what types of kidnapping might be included in its measure; nor does it instruct coders as to how to treat

40. 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.

41. U.S. Dep't of State, 2005 Country Reports on Human Rights Practices: Guatemala (2006), available at <http://www.state.gov/g/drl/rls/hrrpt/2005/61729.htm>.

42. *Id.*

43. *Id.*

44. PROJECT CODING MANUAL, *supra* note 24, at 14. CIRI codebook informs coders not to count "typical" kidnappings as disappearances. *Id.*

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 than does the CIRI.

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;⁴⁵ 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;⁴⁶ the CAR received a 5 on the PTS.

While both countries committed similar numbers of extrajudicial killings in 2006, the Brazilian government was more selective in its application of violence than the government of the CAR, thus justifying CAR's higher score on the PTS. According to the report for Brazil for that year, the majority of the victims killed by security forces were suspected criminals (though forces 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. In fact, because the CIRI does not consider the range of the violence—that is, who is targeted—Brazil appears to be more violent than the CAR.

An additional, closely-related source of difference between the scores is the constraints imposed by the CIRI's component parts. Hundreds of civilians were killed, and scores more were tortured or raped in the CAR in 2006; however, there were no reports of disappearances and the state held few political prisoners.⁴⁷ Given the indiscriminate nature of the killings and torture 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, the state's cumulative CIRI score is only 5. Regardless of how many additional civilians over fifty were killed, and regardless of the

45. U.S. Dep't of State, 2006 Country Reports on Human Rights Practices: Brazil (2007), available at <http://www.state.gov/g/drl/rls/hrrpt/2006/78882.htm>

46. U.S. Dep't of State, 2006 Country Reports on Human Rights Practices, Central African Republic (2007), available at <http://www.state.gov/g/drl/rls/hrrpt/2006/78725.htm>.

47. *Id.*

Table 2
Detailed Example of CIRI and PTS Differences

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

Note: CIRI scores have been inverted for comparison. Herein, a score of "8" for the summed CIRI measure represents the worst possible score while a "0" reflects the absence of each of the types of abuse considered. For the subcategories of abuse, a "0" reflects no cases of that abuse while a "2" reflects the most abusive category.

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). According to the 1994 AI report, Yugoslav authorities tortured or abused thousands, detained hundreds, and killed numerous persons (not to mention the Serbian government-complicit violence in Bosnia that killed many more).⁴⁸ The range of the violence is sufficient for the PTS to consider Yugoslavia as among the most violent countries of 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 1998, the PTS assigned a 5 to Ethiopia while CIRI scored the country an obscenely low (in our view) 3. 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 as did CIRI, the behavior of the Ethiopian government during its first year of conflict with Eritrea seems only moderately abusive.

48. AMNESTY INT'L, AMNESTY INTERNATIONAL REPORT 1995: THE STATE OF THE WORLD'S HUMAN RIGHTS, YUGOSLAVIA (1995).

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."⁴⁹ 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 country's population.

This example raises an additional issue. The PTS considers ethnic cleansing and other mass detention and forcible expulsion 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. In our view, not taking these violations into account in the coding process ignores significant human rights abuses by the state that year.

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 an 8, 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. Thus, for CIRI, because the state engaged in at least fifty of each type of abuse (by the CIRI count), it must be placed in the most abusive category. As aforementioned, the category itself says nothing about the relative severity of the states within it. Thus, once a country passes the threshold number of fifty, it will receive the same score—no matter how great the violence is. In our view, Colombia was vastly more violent than Tajikistan that year, yet both states were 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, like Colombia, and a genocidal state, such as Rwanda in 1994. Still, we feel that despite its claim to greater precision, it is because of this precision that the CIRI is less capable of making these types of distinctions than is the PTS.

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

49. AMNESTY INT'L, AMNESTY INTERNATIONAL REPORT 1999: THE STATE OF THE WORLD'S HUMAN RIGHTS, ETHIOPIA (1999); U.S. Dep't of State, Country Reports on Human Rights Practices for 1998: Ethiopia (1999) available at http://www.state.gov/www/global/human_rights/1998_hrp_report/ethiopia.html.

in dispute or unknown. For instance, the 1997 USDS report on Tajikistan states that there were "a significant number of extrajudicial killings."⁵⁰ In addition, it reports that artillery fire had killed civilians during the country's civil conflict.⁵¹ In neither case are fatality figures provided, nor does the report establish blame. PTS gives the state the benefit of the doubt when ambiguous terminology is used, especially when it is difficult to determine who was responsible for the violence. In addition, the report states that: "A number [sic] disappearances were reported."⁵² However, it goes on to reveal that many of the persons believed to have been abducted were actually forcible state military conscripts.⁵³ While PTS considers this type of abuse in generating a human rights score, forcible conscription is not 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 abused many detainees and arbitrarily detained a large number of persons 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, that there were few actual cases of politically motivated disappearances, and that there was frequent use of torture and arbitrary arrest. We view this as consistent with the description of category 3 presented in Section III.

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 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 (Timor Leste) in 2006 was -0.5. The PTS score for that year was 4, while the CIRI assigned the relatively non-abusive score of 2. The score for CIRI was the result of intermediate scores for killings and torture in that year. According to that year's 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.⁵⁴ The USDS

50. U.S. Dep't of State, Country Reports on Human Rights Practices for 1996: Tajikistan (1997), available at http://www.state.gov/www/global/human_rights/1996_hrp_report/tajikist.html.

51. *Id.*

52. *Id.*

53. *Id.*

54. AMNESTY INT'L, AMNESTY INTERNATIONAL REPORT 2007: THE STATE OF THE WORLD'S HUMAN RIGHTS, TIMOR LESTE (2007), available at <http://archive.amnesty.org/report2007/eng/Regions/Asia-Pacific/Timor-Leste/default.htm>

report confirms much of this statement, though it reports that thirty persons were killed by security forces and other actors.⁵⁵ 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 abuse 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.⁵⁶

VI. CONCLUSION

This research note has attempted to clarify 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.⁵⁷ While this deeper look at the internals of the PTS may seem quite tardy to many users, we believe that something of a reintroduction and clarification is in order. This is especially true given the advent of recent datasets such as the CIRI that offer similar measures of state respect

55. U.S. Dep't of State, 2006 Country Reports on Human Rights Practices, East Timor (2007), available at <http://www.state.gov/g/drl/rls/hrrpt/2006/78772.htm>. Both the Department of State and Amnesty International 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.

56. AI scores more often trend toward category 3 while USDS scores trend toward category 4.

57. See Gibney & Dalton, *The Political Terror Scale*, *supra* note 3, for an earlier discussion.

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 shed light on important differences between the two measures. Most quantitative human rights and political violence scholars have accepted 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.

VII. POSTSCRIPT REMARKS TO CINGRANELLI AND RICHARDS

We thank the editors of *Human Rights Quarterly* for providing space for this exchange between those who produce the Political Terror Scale (PTS) and the Cingranelli and Richards (CIRI) index.

We begin with the issue of the lack of transparency in the PTS. This is not the first time this issue and the absence of a coherent codebook have been brought to our attention. PTS coders have always been instructed to take notes on the abuses for each country, which are used as a basis of comparison in those instances when discrepancies in scores arise. We believe a partial and straightforward fix to the transparency issue raised by David Cingranelli and David Richards is to post these notes on the PTS website.⁵⁸ This will provide more information about how and why we assigned the particular scores that we did for each country. We want to assure the reader that we have never attempted to purposely hide the work that we do, as Cingranelli and Richards seem to suggest. Rather, the decision was based more on the idea that scholars would not find our "field notes" and internal communication all that useful. We welcome the increased emphasis on replicability and transparency in the social science community, and we only regret that we had not made this decision in a more timely way. As always, we welcome suggestions and challenges from other scholars who might disagree with some of the results that we have produced.

58. We plan to start making notes publically available beginning with the release of the 2009 PTS scores in the Fall of 2010. See <http://www.politicalterror scale.org>.

Unfortunately, we fail to see this same spirit of cooperation in Cingranelli's and Richards' article. We believe that both datasets are useful and both have made a significant contribution to the human rights field. However, users need to be aware of the similarities, but also the differences between the two datasets. We do not believe, or would ever claim, that "our" dataset is "more valid" than "theirs." Rather, we attempted to point out key differences and raised what we believe are valid questions about not only the CIRI index's approach, but also about issues in the measurement and conceptualization of human rights. Fighting over which measure is "superior" will do little to improve the field's grasp of the complex and difficult task of quantitatively accounting for state respect for, or abuse of, human rights. Identifying room for improvement in existing measures and striving to address the obvious gaps and limitations in them will contribute to a real aggregation of knowledge.

In that vein, we are singularly disappointed that Cingranelli and Richards made so little attempt to address many of the questions and concerns that we raised in our article. To reiterate, we believe that a mechanical application of their disaggregation scheme can lead to some misleading and even strange results when the individual components are reassembled into a single index. We agree with Cingranelli and Richards that disaggregated measures are important when the empirical questions focus on "types" of abuse. In that case, we would certainly propose that scholars use the CIRI index. Yet, many theories focus on the overall treatment of citizens by their government—that is, they are interested in the conditions, structures, or decision calculus that leads to abuse generally and not to specific kinds of abuse. In these cases, users should be aware of the limitations and (in our view) odd totals arrived at by the aggregated CIRI index. Let us be clear: we are not advocating that abuse is unidimensional; we are simply arguing that the sum of abuses as collapsed in the PTS measure provides a different and more holistic approach to the aggregate levels of abuse committed by a state compared to the CIRI index. Our intention was to be challenging, but ultimately to be helpful. We are not convinced that this has been reciprocated. Both projects use the same source of data and in most instances use of either dataset will result in the same or similar findings. Finally, both projects struggle mightily with the task of trying to make human rights more meaningful and accessible to scholars, practitioners, and governments.

We now turn to several specific issues. The first relates to the use of counts-based categories.⁵⁹ One of the supposed advances of the CIRI index was that it was to provide a more rigorous accounting of human rights prac-

59. David L. Cingranelli & David L. Richards, *The Cingranelli and Richards (CIRI) Human Rights Data Project*, 32 HUM. RTS. Q. 401, 407–08 (2010).

tices than what the PTS offers. Toward that end, the article in which it was introduced specified numeric boundaries for the different categories.⁶⁰ Our criticism is that there is no theoretical justification for these numbers, thus rendering them arbitrary—a view that Cingranelli and Richards now seem to share. However, they explain that: “This scheme was implemented to try to increase intercoder reliability in an era where the source material (the same as used by the PTS) was less systematic than it is at present.”⁶¹ However, they then go on to say that “these numeric thresholds are rarely used to produce scores because usable numeric estimates are seldom provided in our qualitative source material.”⁶² This is correct, and is one of the principal reasons why the PTS has never attempted to rely on specific counts. What is puzzling about Cingranelli’s and Richards’ explanation is that they claim to only infrequently rely on real numbers, but then criticize the PTS coding scheme for its subjectivity. If CIRI coders are not expected to rely on real values, then we assume they are instructed to gauge abuse by evaluating the language used in the reports and by placing a state’s abuse within one of three categories, two of which (“some” and “frequent”) are inherently subjective. In this sense, the coding practice is hardly any less subjective than the one used by the PTS for over twenty-five years. Our point is simply that the CIRI creators can’t have it both ways. The coding is either based on real numbers (which, we are now told, is rarely used), or it is based on a subjective assessment of the frequency with which an abuse occurs.

Another issue relates to the problem of truncation.⁶³ As Cingranelli and Richards point out, “All ordinal scales suffer from this issue, to some extent.”⁶⁴ While we generally agree with this statement, we think that CIRI suffers from an extreme case of this. They go on to say that the producers of the PTS “criticize CIRI for having a scale where a country that tortured fifty-one persons would get the same score as a country that tortured 3,001 persons. That is true of CIRI and it is true of the PTS as well.”⁶⁵ Actually this is not true of the PTS.⁶⁶ Assuming that two countries were roughly of the same size, we see vast differences in the human rights practices between a state where fifty individuals were tortured and another where 500, 5,000, 50,000 (or 3,001) were subject to this odious practice. Most importantly,

60. Cingranelli & Richards, *Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights*, *supra* note 4, at 409–10.

61. Cingranelli & Richards, *The Cingranelli and Richards (CIRI) Human Rights Data Project*, *supra* note 59, at 401.

62. *Id.* at 407.

63. *Id.* at 408.

64. *Id.*

65. *Id.*

66. As a side matter, we did not use the 3,001 figure, but instead, had asked why under the CIRI index countries where 500, 5,000, or even 50,000 individuals were tortured would get the same score as a country where fifty people had been abused in that fashion.

to a large extent the PTS would reflect these differences by assigning these countries different scores; CIRI would not.⁶⁷

Cingranelli and Richards also criticize the PTS for coding countries left untouched by CIRI based on the view that certain states did not have a functioning government, at least for some period of time. This list includes: Sierra Leone, Afghanistan, Somalia, Iraq, and the Democratic Republic of Congo. More than this, the authors warn:

Several studies report results using both the CIRI and PTS indices. The results using the PTS scores always include more observations, but those extra observations result from the PTS project's inclusion of countries with no governments. Because of the different foci of the two projects, using the PTS and CIRI data as mutual robustness checks in empirical research projects is unnecessary and invalid.⁶⁸

While we are deeply aware of the difficulties of attempting to determine the degree of government involvement in violating international human rights standards, we have enormous trouble accepting the principle that because this will not always be clear-cut, the appropriate course of action is to avoid coding these countries altogether. Beyond this, Cingranelli and Richards note that while the PTS has scored Afghanistan and Iraq at level 5 (maximum terror) for the last few years, the CIRI scores, "which only considered the practices of the Afghani and Iraqi governments, indicated less use of state-based terror."⁶⁹ What we do not understand is how and why the Afghani and Iraqi governments—both of which have adopted policies that allow the United States to conduct military affairs in those countries—are thereby exonerated from any and all human rights violations that result from the actions of the United States and allied forces. The score presented by the PTS accounts for abuses committed by agents working at the behest of the body governing the state, as well as those committed by the state proper. We openly admit that these divisions are not always cut and dry, which is also why we account for paramilitary and other pro-government militia abuses. This is an important issue that future data projects may wish to tease out, but we would feel remiss in neglecting the often-significant abuses conducted by

67. In general, according to the PTS coding scheme a country in which fifty persons were killed would in most circumstances not receive a score of 5 but rather a 3 or 4 depending on the extent of other abuses as well as who were the targets of abuse. A state in which 500 persons were killed could possibly receive a score of 4 if the targets were primarily politically active persons rather than apolitical citizens; the state would more likely receive a score of 5 if the abuse was widespread and indiscriminately employed against the populace regardless of its political or social activities. It would be extremely unlikely for a state killing 5,000 or 50,000 persons to receive less than a 5.

68. Cingranelli & Richards, *The Cingranelli and Richards (CIRI) Human Rights Data Project*, *supra* note 59, at 413.

69. *Id.* at 412.

puppet governments and similar entities that do not (for a variety of reasons) meet some formal definition of “government.”

In addition, we are bewildered by the authors’ assertion that using the measures as mutual robustness checks is “unnecessary and invalid.” As repeatedly stated, the measures both code the same types of violations from the same source data. The fine distinction between state practices and conditions might be valid in some general sense, but as we have stated, the PTS has always focused on “state-sponsored political terror.” The CIRI may adopt more stringent definitions in its coding of data, but that does not mean that the two datasets are capturing two completely different phenomena. If they were, the routinely similar results from studies using both measures would seem quite odd. Moreover, our own analysis has demonstrated that the two scores are actually quite similar. And, as we stated, the differences result more from coding procedures, definitions, and CIRI’s arithmetic than from the different “foci” of the projects.

One of the most surprising aspects of the Cingranelli and Richards response is that it is not apparent to us that they truly understand how the PTS operates. We will use two examples. The first is that the authors accuse the PTS of having an *a priori* ordering scheme, and their justification⁷⁰ for this (erroneous) conclusion is that there is no mention of “disappearances” until level 4. They then point out that both Argentina and Azerbaijan had “disappearances” in 2006 but received different scores—as if there were something unusual about this. To be clear, the PTS employs no *a priori* scheme, and there is absolutely nothing sacrosanct about “disappearances” as such. It is unfortunate that there is no mention of disappearances until level 4, but there is simply no meaning (hidden or otherwise) behind this. As we have always readily acknowledged, when it was created the PTS simply adopted the “political terror” scale from the 1980 Freedom House Yearbook. In practice, this means that it is possible that a level 1 country might experience a few disappearances while a level 5 country might not have any. What matters—and what the PTS measures—are the overall human rights practices of a particular country.

What makes this accusation even more puzzling is that Cingranelli and Richards reiterate an argument they have made before that human rights violations occur in certain stages (which certainly sounds like an *a priori* ordering to us).⁷¹ Yet, rather than assigning additional weight to certain violations as one advances up this purported scale, the CIRI index treats all four categories—political imprisonment, disappearances, summary executions, and torture—exactly the same. In our view, it is misleading to treat a political imprisonment the same as a political killing.

70. *Id.* at 418.

71. *Id.*

The second example relates to “violent expulsions.” The authors are indeed correct that there is no mention of such practices in the PTS description—and neither is there any mention of such practices under the CIRI index. Yet, what is most interesting is how differently the two projects treat this phenomenon. As we state in our article, the PTS seeks to reflect this on the grounds that a violent expulsion constitutes a serious violation of an individual’s physical integrity rights. In this sense, it is quite similar to arbitrary imprisonment in that it is typically accompanied by physical and psychological deprivation and torment and often constitutes inhumane or cruel treatment. In contrast, CIRI does not consider this treatment unless it specifically fits into one of the defined categories. The authors do acknowledge the PTS coding decision on this issue as being “reasonable,” but they then go on to accuse the PTS project of a lack of transparency. In our view transparency is not the issue as much as the rigid and narrow application of definitions that ignores broad classes of clearly abusive behavior.

Another difference relates to the size of a country’s population. As we point out in our article, the PTS factors in a country’s size while CIRI does not. The idea, quite simply, is that a thousand political prisoners in China means something much different than a thousand political prisoners in a much smaller country, such as Cuba. We will acknowledge that we have not approached this issue as systematically as we might have, but what we find puzzling is the claim by Cingranelli and Richards that factoring in a country’s population size would be a “mistake” because it “introduces too much subjectivity into the coding process.”⁷² We stand by our assertion that the relative scope of violations within a state matters. To return to the comparisons we used earlier, the CIRI scale consistently places populous countries such as China and India in the worst categories of abuse. Although there are certainly high levels of human rights violations in both states, it is simply not accurate to place China and India in the same category as the likes of the Democratic Republic of Congo and the Sudan. Yet according to CIRI’s count this is exactly what occurs. We direct readers to the other country comparisons in the main article.

After chiding us for including population size because of its “subjectivity”—in what way is a country’s population size subjective?—the authors then go on to lecture us for giving larger countries a free ride, but also for missing the essence of human rights:

International law, according to any reading with which we are familiar, does not give large countries special dispensation on violations because of their population size. No matter a country’s population, the standard of judgment

72. *Id.* at 420.

is that *no one* is supposed to have his or her physical integrity violated. To the extent that the PTS does not use law as its standard of behavior, it drifts from being a standards-based rating towards being a relative ranking of countries. Further, we must illuminate an inherent ethical dilemma in this method: using a sliding scale based on population means that some deaths count more than do others. This is a violation of the basic principle of dignity that each life is of equal worth.⁷³

We simply do not understand the charge that the PTS is not based on international human rights law. What are disappearances, summary executions, and so on but violations of international human rights standards? And we assure the two authors that while their reflections on the meaning of human rights were most assuredly well intended, these comments were simply not needed. The larger point is that the issue at hand has little to do with the equal worth of all human beings, which serves as the core principle upon which the entirety of human rights is based. Rather, it is about trying to create some meaningful standards by which state behavior can be tested. In our view, employing an events-based approach that does not factor in the size of a country's population does not meet this standard.

Our final point relates to the issue of aggregation. In our view, the most significant contribution that the CIRI index has made is the manner in which it disaggregates a state's human rights practices. Because of this, we now know much more about the manner in which human rights violations take place in each country. However, it is also important to point out that the CIRI index tells us nothing about who these practices are directed towards. Like putting Humpty Dumpty back together, we think that there can be serious problems when these disaggregated measures are then aggregated. This simple but important point is one that Cingranelli and Richards repeatedly ignore.

73. *Id.* at 421.