

**Differences in degree or differences in kind?
Re-conceptualizing State Repression**

Ragnhild Nordås (rmordas@umich.edu) and Charles Crabtree

Abstract:

Cross-national studies of physical integrity violations generally conceptualize state repression as a scalar phenomenon where the same drivers are important for all levels of repression (i.e., differences in degree). This assumption has led to theorizing, measurement, modeling, and findings that confirm this view as well as diverse policy implications regarding the reduction of state repression. However, we argue that the conclusions derived from the scalar assumption are a function of how that particular assumption has guided research. We challenge the scalar assumption and conceptualize repression instead as a discrete level phenomenon where different variables impact distinct levels of repression (i.e., differences in kind). By estimating a canonical model from the state repression literature using Random Forest classifiers, we find that key conclusions based on the scalar assumption are deceptive, and that theorizing and modeling repression as discrete levels should improve our understanding. Specifically, we demonstrate that *no* independent variable consistently predicts outcomes across the standard repression measure of the Political Terror Scale (PTS) and some variables are extremely important for particular levels. We conclude that future work in this field should explore additional theoretical mechanisms that link country-level factors to specific repertoires of human rights abuses, and model repression by discretizing human rights outcomes.

Keywords: State repression, physical integrity violations, political terror, discrete outcomes

Paper for presentation at Michigan State University, 19 October 2018

Acknowledgements: We would like to thank Christian Davenport, Håvard Strand, and Reed Wood for extremely valuable feedback and support.

Introduction

The last three decades has seen hundreds of articles and a few books published seeking to shed light on what influences repressive behavior through systematic cross-national investigation (see, e.g., Apodaca 2001; Bagozzi and Berliner 2016; Bueno de Mesquita et al. 2005; Cingranelli and Filippov 2010; Cingranelli and Richards 1999; Conrad and Moore 2010; Conrad et al. 2013; Davenport 1995, 1996, 1997, 1999, 2007a, 2007b; Davenport and Armstrong 2004; de Soysa and Nordås 2007; Fariss 2014; Fein 1995; Hafner-Burton 2005; Hafner-Burton and Tsutsui 2005; Hathaway 2002; Henderson 1991; Hill and Jones 2014; Keith 1999; Keith, Tate, and Poe 2009; Mitchell and McCormick 1988; Neumeyer 2005; Nordås and Davenport 2013; Poe and Tate 1994; Poe, Tate, and Keith 1999; Richards, Gelleny, and Sacko 2001; Richards, Webb and Clay 2015; Zanger 2000). Motivated by a desire to understand and stop such state violence, this literature has been built on the assumption that any insights gained about what influences repression (at any level of severity) are useful for constructing policy interventions to limit states' repressive behavior.

This notion is evident in how quantitative scholars have theorized, measured, and modeled repression as being *scalar* where different levels of repression levels are influenced by the same variables using regression models designed for specific dependent variables. Such an orientation is important to recognize, because it means that the work is guided by an assumption that diverse explanatory factors are equally informative about lower *levels* of state repression as they are about *higher* levels. In this paradigm, numerous variables have proven to be important for influencing repression. For example, regime type, civil conflict, and youth bulges have all been consistently found to be both statistically significant and substantively important (e.g. Hill and Jones 2014).

In contrast to the scalar view, we propose a reconceptualization of state repression which we refer to as “discrete”, where different explanatory variables influence different repression levels. Theoretically, we argue that the previously applied cost-benefit argument undergirding most of the repression literature has failed to consider how perceived costs and benefits are likely to vary across different levels of repression, making different explanatory factors vary in their relevance at predictors of different repression levels. In short, we cannot assume that repression is only a matter of difference in degree, but that it might be better understood as difference in kind. Such a re-orientation sets us apart from the scalar paradigm in important ways, because it means that diverse explanatory factors are potentially informative about different levels of repression – in other words, suggesting that some variables explain some levels of repression whereas other variables explain other levels, either because of how a uniform scale collapses critical dimensions and aggregates differences in kind (or form), or because the cost-benefit calculus looks dramatically different at various levels of repression, hence that the theoretical argument needs development. Whether this is a measurement or theory problem first and foremost (or, we would argue, it is both), a solution is to re-conceptualize and model state repression as discrete levels.

We demonstrate that re-conceptualizing state repression by considering the discrete approach leads to new empirical insights as well as a rethinking of fundamental policy implications of previous research. It also opens up new avenues for theorizing about repression. On the one hand, a shift in theoretical orientation and modeling from scalar to discrete approaches will not only provide greater insights into which levels of repression are consistently influenced by different explanatory variables but also will allow us to address which repressive levels are not likely impacted. This has relevance for not only research on human rights violation but also public policy research regarding what influences as well as stops relevant behavior. Given the differences

that we find exist with regard to what matters for different levels of repressive behavior (e.g., between protest policing on the lower end and genocide at the higher end) it makes sense to consider them individually. On the other hand, the discrete paradigm also represents a superior approach to social science. For example, the scalar approach hides the discovery of discrete relationships but the discrete approach allows for the discovery of scalar relationships. The discrete approach thus leaves as an empirical question what the relationship is between different variables and repression without presupposing scalar properties that might not hold.

Guided by the discrete conceptualization and employing a more flexible modeling strategy fitting this reorientation, we focus our empirical investigation on the predominant measure for state repression -- the Political Terror Scale or PTS (Wood and Gibney 2010).¹ By allowing for the possibility that the individual levels of PTS might be substantially different and driven by different processes, we examine and reveal new empirical findings as well as theoretical puzzles. For example, we find that key theoretical variables from the existing literature differ in the degree to which they can predict discrete levels of human rights abuses (e.g., executive constraints predict levels 1 and 3, trade is a predictor of levels 2 and 4 on the PTS in particular; and GDP per capita predicts level 1 well, and has some importance in predicting levels 2-4, but not level 5). Equally as important, no independent variable can consistently predict the full range of human rights outcomes in the PTS. Finally, results disclose that many variables in existing models fare poorly at predicting specific outcomes. In particular, if we want to predict the worst level of state abuses (level 5), we find that only *one* variable has strong predictive power, namely civil war, and none have a negative impact -- which suggests that nothing predicts well how we prevent this level.

¹ An alternative measure that is also often used is the Cingranelli and Richards (1999, 2010) index of governments respect for physical integrity rights. However, as the PTS is arguably the most frequently used measure, we focus our empirical investigation here.

These findings are noteworthy in that they reveal how the prevailing scalar assumption has hindered our understanding of what drives repressive behavior, and it reveals new empirical results at odds with the conclusions drawn from prior scholarship. Taken together, our conceptual discussion, measurement and empirical findings provide justification for discretizing existing human rights scales in future work. This approach should also open up important new avenues for research on state repression. In particular, considering repression levels as discrete outcomes (rather than scalar) forces, or at least encourages, researchers to think more carefully about the theoretical mechanisms that link country-level factors to specific categories (or repertoires) of human rights abuses.

The paper proceeds in four sections and a conclusion. First, we briefly lay out the genesis of the cross-national empirical study of state repression and the scalar assumption underlying most of this work. We then present our discrete approach and how it opens up different theoretical as well as empirical lines of inquiry. Next, we discuss how the discrete approach leads to different empirical choices, and outline how Random Forest classifiers facilitate our investigation of repression as discrete outcomes. We then present our empirical findings which challenge established understandings of what drives repression. Last, we outline how new theoretical puzzles emerge from these results.

Repression research

The beginning of the systematic study of state repression can be traced to a series of historical evaluations of diverse levels of repressive behavior in the 1960s and 1970s (e.g. Dallin & Breslauer 1970; Walter 1969; see Davenport 2007 for a review). This work essentially described the basic phenomenon involved (i.e., the repressive behavior of interest) and provided detailed accountings of who did what to whom, where and when, with only some discussion regarding why.

Highlighted within this earlier work is the role played by political institutions – especially authoritarian and totalitarian ones. For example, scholars maintained that in such regimes lacked checks on leaders and security force agents who wished to engage in repressive behavior, and that this therefore prompted repressive action. In contrast, democracies have diverse checks on authorities and security force agents as well as diverse mechanisms of control beyond coercion and force (i.e., channeling behavioral challenges into pre-existing networks and processes). This would result in democracies decreasing their use of repression. At this early stage, however, this difference was not explicitly and systematically investigated.

More theoretical explanation for why we see repression began to take shape in the 1980s. Influenced by trends within the broader literature on conflict and violence, researchers began to move away from a structural argument (repression as determined by institutional design) toward something that would be classified as rational choice. With that, repressive behavior resulted not from a particular institutional configuration, but from some leader's evaluation of political, economic and demographic context, placed into particular categories (i.e., benefits and costs) and then evaluated with regard to their ability to maximize utility. This cost-benefit (but particularly emphasizing costs) has since come to dominate theorizing on state repression. The theoretical argument maintains that political authorities evaluate the context as to whether it favors or disfavors the use of state repression, and that authorities adjust their use of repressive behavior upward or downward in accordance to the relevant context. As the context and resulting decision calculus favored repressive behavior (particularly contexts of threats to the regime), government coercion and force would increase.

Differences in Degree: The Scalar Approach

The most important implication of the cost-benefit theories embedded in current research, is that repression is conceived of as being a uniform scalar phenomenon; and the government views the components of the decision calculus as being relevant for the full scale. Benefits are equally valued across tactical choices, as are costs. A largely implicit consequences of this approach is the idea that an individual variable has an impact on repression across the full scale of repressive behavior – we are trying to explain differences in degree, and if we accept that proposition, the same explanations should work in similar ways across this uniform scale. Thus, if a particular variable has an impact it is believed that increasing or decreasing that variable would have implications for moving repression upwards or downwards across all its values.

Figure 1 (below) illustrates the scalar assumption visually, where each variable (exemplified with A, B, C, D, and E as placeholder examples) is assumed to have a significant predictive power for each of the levels (1-5).

Figure 1. Illustration of the expectations built into the scalar assumption

Types/levels Variables	1	2	3	4	5
A					
B					
C					
D					
E					

Within this framework, researchers have engaged in the task of operationalizing the different components of the theoretical cost-benefit model (with varying degrees of success). This involved factors such as the domestic political system (e.g., level of democracy, type of democracy, characteristics of democracy [holding elections, parties, executive constraints, media, judicial independence]), economic relations and development [e.g., GNP, trade, economic

sanctions] and diverse demographic characteristics of the population [e.g., youth bulges, size of the population]).

Equally as important is the measurement of state repression. Not surprisingly, the scalar theorizing of repression and the measurement of state repression as a scalar phenomenon evolves symbiotically. In the area of repression and human rights violation, the discussion of scalarity was initiated in the 1980s (see Stohl et al. 1986; Stohl and Lopez 1986). Here, researchers discussed what should be included in a repression scale, whether such a scale was possible and useful, what sources would be best, what the different forms of repression were to be included, what scaling technique should be used, and what were the underlying dimensions that researchers would be identifying.² These discussions were fundamental to the development of a repression measure³, but largely died down after the introduction of key physical integrity rights measures, most notably the PTS dataset.⁴ Hence, rather than a unified discussion of general measurement of state

² The discussion rests on a notion that there is an underlying dimension that lay beneath different manifestations of political conflict that are believed to be connected in a way that necessitates that they be conceived of and represented together in a single indicator, and that we can determine what this scale is identified by and seek explanations for where on the scale a case ends up. This fundamental connection from how the world is perceived to operate and what path that leads scholar down in terms the intuitive empirical choices following from that conceptualization, is what Tilly (2002) has referred to as the theory of measurement.

³ Scalarity has also long been pursued in a range of social science research, such as the measurement of political democracy (Bollen 1980), development (Booysen 2002), culture (Hofstede 2003), ethnic conflict (Vanhanen 1999), and property rights (Knack and Keefer 1995) – often subjecting the measures to factor analysis to identify underlying dimensions (Bollen 1980) that are subsequently labeled and analyzed. For many of these other research areas, however (in particular democracy and development) the debate about the scales applied have arguably been more comprehensive and rich.

⁴ An exception is Landman (2004) and Landman and Carvalho (2010) discussing how human rights can be measured in principle, in practice, and as outcomes of government policy, and discussing various approaches to measuring these dimensions, including “the coding of formal legal documents, events-based, standards-based, and survey-based data, as well as aggregate indicators that serve as indirect measures of rights protection” (Landman 2004: 906). Landman identifies the PTS as a standards-based scale. There have also been some work discussing how the PTS and CIRI

repression of human rights, much of the cross-national comparative literature on repression coalesced around using the PTS, and to conceptualize and model repression as scalar.⁵

PTS has a long history of use as an indicator of state repression and despite recent advancements in human rights measurement (e.g., Fariss 2014), it remains a popular choice. The PTS measures levels of political violence and terror that a country experiences in a particular year based on a 5-level “terror scale”. The scale is presented as a continuum of human rights practices, but is constructed from aggregation of three different underlying dimensions of repression: 1) the type of violence being carried out by the state (specifically imprisonment, torture, and killing); 2) the intensity of the abuses, or the frequency or number of instances of a given type of abuse in any given year; and lastly, 3) the portion of the population targeted for abuse, which is referred to as the range of political terror (Wood and Gibney 2010). These three dimensions are then combined to a composite scale. The levels of the PTS scale are shown in Table 1, below.

differ (see Wood and Gibney 2010; Cingranelli and Richards 2010), and before that there was also some work on how the two main sources of data for these measures, the US State Department and Amnesty International introduce systematic biases that produce different views of the situation of human rights at different points in time and in different regions (Poe et al. 2001).

⁵ In addition, separate strands developed focusing on particular outcomes that form part of repressive repertoires (e.g. ethnic discrimination, indiscriminate violence, protest policing, counter-insurgency, mass atrocities, genocides) (e.g. Harff 2003; Valentino et al. 2004). These relatively disparate literatures have each produced insights about the variation in the particular forms of violence or coercion seen in isolation. Although some attempts have been made to discuss “whether distinctions within the concept [of repression] are productive and/or forms of repression are directly comparable” (Earl 2011, 261), the findings are rarely unified through empirical comparison across different types of state behaviors.

Table 1. The Political Terror Scale

Level	Description
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.
2	There is a limited amount of imprisonment for nonviolent political activity. However, few persons are affected, torture and beatings are exceptional. Political murder is rare.
3	There is extensive political imprisonment, or a recent history of such imprisonment. Execution or other political murders and brutality may be common. Unlimited detention, with or without a trial, for political views is accepted.
4	Civil and political rights violations have expanded to large numbers of the population. Murders, disappearances, and torture are a common part of life. In spite of its generality, on this level terror affects those who interest themselves in politics or ideas.
5	Terror has expanded 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.

Many studies have been published using the PTS measure as the dependent variable, and modeling variation in this measure relying on a scalar assumption (e.g. Davenport and Armstrong 2004; de Soysa and Nordås 2007; Neumayer 2005; Nordås and Davenport 2013; Poe and Tate 1994; Poe et al 1999; see PTS for complete listing). In this work, we can clearly see how the expectations and results are presented in language that indicates a scalar logic. In the seminal work by Poe and Tate (1994), they largely established the set of variables that researchers would continue to investigate in some manner for the next 24 years. For example, they find that “democracy was shown to be associated with a decreased incidence in repression” (p. 866), indicating that they conceive of repression as a uniform scalar concept which particular variables uniformly affect – producing increases or decreases. Later, Poe et al. (1999) stated that they “expect repression to be less extensive in democracies than in other types of government” (p. 293), indicating that the dependent variable captures “more” or “less” extensive repression (ibid.) and that democracy has a constant effect on this scale; or that different variables lead to “greater human

rights abuse” (p. 291) or “to less abuse of personal integrity rights” (p. 310).⁶ Poe and Tate (1994) find that “economic standing is negatively (...) related to regimes' propensities to abuse of personal integrity rights”, again indicating that there is a scalar logic, that economic standing has a uniform impact on this scale, which can be understood in terms of increases or decreases in repression or, as in this example, as different propensities to engage in abuse of personal integrity rights.

The scalar literature has gradually coalesced around a set of variables capturing i.a. political system type, economic development, and population size, as well as the occurrence of intrastate armed conflict/civil war as a driver of repressive acts by states (e.g. Hibbs 1973; Henderson 1993; Poe and Tate 1994; Davenport 1995; 2007a,b; Krain 1997, 2005; Richards 1999; Davenport and Armstrong 2004; Hafner-Burton 2005a,b, 2008; Richards and Gelleny 2007; Wood 2008; Abouharb and Cingranelli 2009; Carey 2009). In a recent re-analysis of the literature’s primarily empirical claims, Hill and Jones (2014) identified which of the indicators used to capture the different conditions assumed to drive repression perform the best. They find that civil conflict is the measure that adds the most predictive power to a baseline model of log of GDP and population, while once civil war is included in the baseline model, the most powerful indicators are youth bulges, political competition, judicial independence, common law legal systems, executive constraints, and fair trial provisions (Hill and Jones 2014, p. 672). These core variables identified as the most consistent predictors are in line with the general observation that domestic factors outperform or provide more consistent support for levels of state repression than international factors (ibid.; Nordås and Davenport 2013). The standard list of variables to include in a repression model now therefore usually includes measures of GDP per capita, population size, civil and international war, and democratic political institutions. As we discussed above, the supposed effect

⁶ Similar wording includes talking about “determinants of personal integrity abuse” which also implies that repression is conceived of as scalar, even if this is not explicitly stated (ibid. 310).

of these factors is assumed to be uniform across the PTS, and the empirical results are discussed with the notion that these variables are important across the entire scale as a stated or implied tenet.

This observation does not only apply to researchers adopting a research approach. Even when researchers use ordered outcome models, they typically discuss their results from these models in such a way that the effect of independent variables is presumed constant across levels of the dependent variable. For example, Neumeyer (2005) clearly suggests that this research area runs into measurement and modeling problems, but presents the analysis of impact of international human rights treaties on PTS as identifying how countries fare on a scale of “human rights performance”. Similarly, de Soysa and Nordas (2007) also refer to findings that particular country-level indicators related to religious makeup are associated with “lower levels of repression” (p. 936) than countries with other traits. An issue here is that researchers do not present predicted probabilities (or other quantities of interest) across categories. This means that, in effect, there is still a *de facto* scalar assumption. As we show later, a fruitful alternative to this approach is to consider each category as a discrete (binary) state.

Figure X. Summarizing main variables in literature (considering adding this) – NOT COMPLETE

Variables	Theorized Impact	Estimation Strategy
	(positive/negative)	(measure)
Regime type (democracy)	Negative	
Youth bulge	Positive	
Civil war	Positive	
Trade	Negative	
Etc.		

The Discrete Approach

In contrast to the scalar paradigm, we suggest that a discrete approach might be more useful for improving not only understanding but efforts to curb relevant behavior. *What do we mean by discrete?* Effectively, this means that we should not think of repression as being a scale – at least not by definition, precluding alternatives. Theoretically, this means we are proposing that it is possible that the different levels of repression as in fact differences in kind more so than degree. If so, different levels should be driven by different factors or by unique/distinct combinations of variables, which the scalar assumption conceals. Again, if this holds, we would have to reconsider or at a minimum develop the cost-benefit argument in existing studies, or look for other theoretical explanations and associated measures that might successfully predict different levels of repression. Findings supporting a discrete conceptualization would also suggest that alternative scales or dimensionality could be derived and tested. For example, we can evaluate repressive behavior varying in scope, lethality and victim [population](#)[NR1].

Why do we expect repression to be better understood as discrete levels? To illustrate this, we can take our cue from the literature concerning selective, indiscriminate and collective violence (e.g. Kalyvas 2006; Valentino et al. 2004). This research is very clear about the fact that although all concern a form of political violence, the different forms are influenced by very different factors. For example, selective violence is believed to be employed when governments are generally strong and efficient in identifying behavioral challengers. In contrast, indiscriminate violence is used when governments are weak and incapable of identifying challengers. A third circumstance exists when collective violence is used. In this situations, governments are capable of identifying who they wish to target and they are capable of employing violent activity in a consistent and widespread manner. Underlying this research is a concern with capacity but overlaid with this is a concern with objectives. The most important thing about this work however is the fact that while

one can discuss the various forms of political violence at the same time, their use appears to emerge from very distinct explanations and explanatory factors. This prompts us to look for influences in a specific way that is more in line with the discrete approach outlined above.

The crux of our argument revolves around what we call the *questionable presumption of equitable consideration*. This refers to the fact that generally researchers maintain that political authorities consider all explanatory factors when they are deciding whether to engage in state repression. This we believe is highly unlikely. One illustrative example of the appropriateness of equitably considering all explanatory factors can be when we consider that state repression can vary in the degree of (political) targeting involved (i.e., the degree to which authorities select targets for political reasons), which emerges as one of the oldest concerns in the area of state repression. In the post-World War II environment, there was a significant amount of attention given to evaluations of totalitarian and authoritarian political systems related to state repression (e.g., Dallin and Breslauer 1970; Linz 1985). According to this work, more closed political systems tended to perceive greater amounts of threat from civil society and there was also an attempt made to eliminate thinking that was deemed challenging to the political, economic and social ideas associated with these regimes – which at the time were communist or socialist. Lacking economic and normative mechanisms of influence, totalitarian and authoritarian governments tended to rely on repressive action. Contrary, as individuals and groups have an ability to “participate”, they are less likely to substantively challenge the government and thus they are less likely to be targeted by political authorities for political reasons. A hypothesis to emerge from this literature was that more open political systems decrease political targeting. If this logic holds, we expect to see variables that capture political openness as better able to predict discrete outcomes with more political targeting (e.g. PTS levels 2, 3 and 4) than situations in which

repression is largely indiscriminate (level 5), or that a particular concatenation of discrete outcomes could be predicted by set political regime factors. A finding in line with this expectation about differential effects for different discrete outcomes would also challenge the scalar assumption by demonstrating differential effects of key explanatory variables across discrete outcomes on the dependent variable.

International shaming is another factor that has been found to affect repression when viewed in the scalar perspective but which might be better evaluated in the discrete framework. For instance, Murdie and Davis (2012) focus on how INGOs are shaming governments, and how this influences state behavior. Others support this finding as well (e.g., Hill and Jones 2014). However, theoretically, shaming should presumably impact on more visible forms of repression in particular. This means that high levels of lethality (or higher amount or severity of violence employed by repressive agents) which increases the visibility of the repressive tactic should be most readily influenced by external actors shaming strategies. Which level of lethality the state chooses to engage in as part of their repressive repertoire could therefore be a function of a regime's sensitivity to international reputation costs and their capacity to engage in effective covert repression. If this logic is operating, we might expect to see shaming to be more likely to prevent levels 4 and 5 (outcomes involving high visibility and lethality) than having an effect at other levels (1-3). Again, however, findings that suggest such differential effects would also indicate that the scalar assumption is not fitting.

Theorizing and conceptualizing state repression as discrete outcomes is useful in that it opens up discussion about repressive repertoires and theorizing about discrete outcomes such as the examples above help illustrate. For example, the different dimensions of the PTS (i.e. forms, lethality and scope of repression) could be theorized with attention to the tensions that might exist

between them, or we might theorize what is needed to attain particular outcomes, level 1 vs. 3 vs. 5 etc. Unlike in a scalar approach, governments are not perceived as simply deciding to repress or not and at what “level”. Rather, governments are deciding what type of repertoire (i.e., combination of dimensions/tactics) to engage in and against whom, which contributes to producing different outcome levels of repression. This involves making numerous decisions at once: what is the political nature of the targeting, who is to be targeted (scope) and how much violence or other coercive measures (and of what kind) should be used. Although states are likely to respond to threats with repression (Davenport 2007), it is necessary to also take into account how threats can be diffuse or specific, and who precisely is challenging authorities (e.g., Davenport et al. 2011; Nordås & Davenport 2013). Individual and in combination, these decisions processes could determine which discrete level of PTS we observe. Challenging the scalar assumption by conceptualizing discrete outcomes is a first step towards taking these decision-making processes into account, more fully considering variation across repressive repertoires, and testing them cross-nationally.

Finally, this approach also answers an important call to begin to assess precisely how and when diverse forms of political violence should be considered together. In line with Verdeja (2012: 82): we believe that we should seek

the conditions under which it is likely to occur, the small-scale processes (or, in Charles Tilly' s terms, “ mechanisms”) of violent escalation and de-escalation at local levels, and the ways in which these processes are shaped by, connect to, reinforce, accelerate, and impede higher-level processes of violence. We should explain in other words, variability in violent outcomes.

What discrete might look like empirically

The discrete conceptualization can manifest in different ways. A strong version of the discrete approach (which we refer to as “isolates”) would mean that there is no variable that wields an influence across all types or levels of repression. In this view, we should instead consider which variables are likely to produce particular levels of state repression, and test such arguments empirically by focusing on levels as discrete outcomes rather than as necessarily part of a scale. A less strong version (“overlapping across levels”) would maintain that there are a few variables that wield an influence across types, but be geared towards identifying clusters of variables that predict particular outcomes but not others, or identifying particular clusters of outcomes and what variables help explain their occurrence. There is reason to believe i.e. that some variables will be more effective at predicting shifts in the lower end of the political terror scale (what makes a country transition from general respect of human rights to initiating some restrictions), whereas others might be more effective in predicting other shifts on the scale (such as the transition from a level 4 on PTS to the highest level of 5 where repression is without limits and population wide). A third possible expectation from the discrete logic is that there are boundaries of repression level which particular factors are able to predict (“overlapping within levels”). This maintains that the repression levels are ordered, and that it might make sense to talk of lower and higher levels of repression, but the discrete approach does not force a scalar logic on the data, but leaves it as an empirical question.

The discrete way of seeing and modeling state repression, and the three varieties of the discreteness, is illustrated graphically in Fig. 2 (below). Here, the different theoretical possibilities in the discrete approach are represented, with the explanatory variables (A, B, C, D, and E) as placeholder examples. Figure 3 depicts a hybrid of scalar and discrete, where one or two variables (here A and B) predict across levels, and other variables do not.

Figure 2. Examples of possible alternative expectations when treating repression as discrete outcomes

“Isolates”

Types/levels Variables	1	2	3	4	5
A					
B					
C					
D					
E					

“Overlapping across levels”

Types/levels Variables	1	2	3	4	5
A					
B					
C					
D					
E					

“Overlapping within levels”

Types/levels Variables	1	2	3	4	5
A					
B					
C					
D					
E					

Note: Shaded squares designate significant predictive power.

Figure 3. Example of hybrid of scalar and discrete

Types/levels Variables	1	2	3	4	5
A					
B					
C					
D					
E					

Evaluating scalar vs. discrete approaches

The standard way researchers have conceptualized and tested theories about state repression of physical integrity rights is by conducting null hypothesis significance tests (NHSTs) for one or more covariates that measure theoretically relevant concepts (Hill and Jones 2014). This approach to statistical testing has been criticized (e.g. Gill 1999; Hill and Jones, 2014), for various reasons. One that is most important for this paper is that the NHST approach alone does not necessarily reveal which variables, if any, predict outcomes. Instead, the predictive performance of variables and the overall model should be assessed directly. A key issue with previous approaches is that the modeling (and measures used) deliberately or implicitly assume that repression (or the way it is measured) is appropriate or useful to conceive of as a one-dimensional concept, and reasonably captured on an ordinal scale with (often) the steps of the scale equidistant to each other.

Data and Methods

To evaluate the appropriateness of the scalar approach versus our alternative approach of considering repression levels as discrete outcomes, we relax the scalar assumption undergirding existing scholarship to better gauge whether the restrictions imposed by these assumptions are limiting our understanding of drivers of repression. We do this recognizing the empirical and theoretical advantages of treating the levels of repression from PTS as discrete outcomes. It has been assumed that the key explanatory variables proposed by existing studies are equally valid predictors across the terror scale (Figure 1). By shifting to a discrete focus we can unpack this assumption and determine whether, in fact, particular variables are able to predict particular outcomes and a different set of variables predict other outcomes, in line with the different alternative expectations presented in Figure 2 (above) or a hybrid (Figure 3).

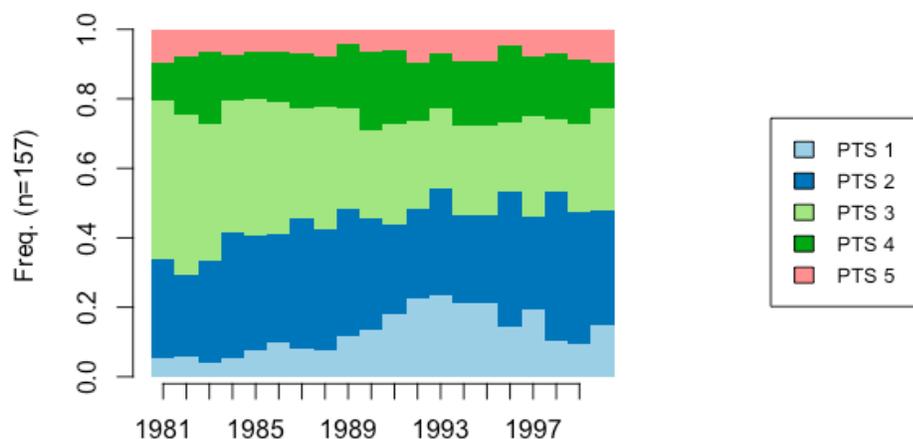
To investigate the determinants of discrete (binary) human rights outcomes, we construct a dataset of 2,224 observations, covering 157 countries from 1981-1999.⁷ Our outcome measure is adapted from PTS (Wood and Gibney 2010). As discussed, the PTS measure ranges from 1-5, with higher values indicating worse respect for human rights. To study repression as discrete outcomes, we transform the ordinal indicator into a series of dummy variables. This gives us five discrete indicators (or outcomes), one for each level of the scale: *PTS 5*, *PTS 4*, *PTS 3*, *PTS 2*, and *PTS 1*.⁸

Since researchers have not previously analyzed these outcomes as discrete states, we visualize them in Figure 3, to provides a sense of the how common these outcomes are over time. Focusing on the extremes of the PTS scale, we see that the number of *PTS 5* countries has remained relatively steady over time, while the number of *PTS 1* countries has slightly increased. The largest movement across our discrete outcomes appears to be from *PTS 2* countries changing to *PTS 3* countries over time.

⁷ We selected this sample to ease comparison with the results reported in Hill and Jones (2016). Our primary findings are robust if we use a large sample of countries and a longer window of time.

⁸ We include kernel density plots for our five new outcomes in Appendix A.

Figure 3. Distribution of PTS Categories Across Time



Note: Figure 3 displays the observed frequency of our five outcome measures (i.e. PTS 1, PTS 5, PTS 2, PTS 3, PTS 4, PTS 5) from 1981 to 1999.

To best compare to existing studies, we choose to include the broad array of different explanatory variables identified in the literature and used in the meta-analysis of Hill and Jones (2016). Table 2 (below) presents the measures used and their original sources.⁹ These factors capture broad country-year differences in history, demographics, size, wealth, conflict history, and political institutions. Individually, these variables often appear in ‘standard model’ specifications of human rights abuses. Taken together, they constitute nearly the census of variables that the theoretical and empirical literatures on repression have identified as the primary determinants of state repression.

⁹ We use the data as provided by Hill and Jones (2016). This means that it incorporates their approach to dealing with missing values. As they describe in their online appendix, they address missingness by using the random indicator method for numeric covariates and random forests for all binary, ordinal, and categorical variables. Interested readers should refer to their appendix for more details.

Since we have discrete (binary) outcome measures, we might estimate the effect of these variables using a logit or probit model. However, advances in machine learning though provide scholars with several other (sometimes better) options. We use one of these algorithms, a Random Forests model, to determine the predictors of individual PTS levels. Since our outcome measures are binary, we estimate a series of Random Forests classifiers.

Table 2. Data Used

Measure	Source
PTS	Wood and Gibney (2010)
Population Size	Gleditsch (2002)
Youth Population	Urdal (2006)
GDP Per Capita	Gleditsch (2002)
Oil Revenue	Ross (2006)
Civil War	UCDP/PRIO armed conflict
Interstate War	UCDP/PRIO armed conflict
Democracy	Polity IV
Military Regime	Database of Political Institutions
Left/Right Regime	Database of Political Institutions
<i>De facto</i> Judicial Independence	CIRI
Constitutional Provisions	Keith, Tate, and Poe (2009)
Common Law System	Mitchell, Ring, and Spellman (2013)
Trade Openness	World Bank
Foreign Direct Investment	World Bank
Structural Adjustment (WB and IMF)	Abouharb and Cingranelli (2007)
PTA Agreement w/ Human Rights Clause	Spilker and Bohmelt (2012)
INGO Presence	Hafner-Burton and Tsutsui (2005)
INGO Shaming	Ron, Ramos, and Rodgers (2005)
Western Media Shaming	Ron, Ramos, and Rodgers (2005)
HRO Shaming	Murdie and Davis (2012)
ICCPR Ratification	Jones (2014)
CAT Ratification	Jones (2014)

Note: In a set of alternative models, we use the regime type measure provided by Davenport and Armstrong (2004) instead of the democracy measure from Polity IV. We obtain similar results in those models.

The primary reason we use Random Forests models instead of (generalized) linear model is that the former provide an established, validated metric of variable importance (Strobl et al 2008). If we were to estimate a logit or probit model, we might classify a variable as important if it is statistically significant at some alpha level (e.g., 0.05), if its inclusion in a model improves some goodness-of-fit statistic (e.g., R^2 or AIC), or if adding it to the model reduces root-mean-square-error. The problem here is that researchers still disagree about which one of these criteria is the best at measuring variable importance (Gromping 2012). Random Forests models, in contrast, provide a validated measure, which we describe in greater detail below, that captures the predictive importance of variables. In many cases, this can be a more informative “measure of substantive importance than, for example, statistical significance tests” (Jones and Linder 2016, 12-13) or other statistics provided in linear model output.¹⁰

The second reason to use Random Forests classifiers instead of traditional statistical tools, such as a logit or probit models, is that we do not have strong *a priori* theoretical expectations about the generative structure of the data. While the repression literature provides guidance about what variables we should include in our models, it provides less clear instruction about the functional form of variables that we include in the model or how these variables interact. Without knowing those things, we would likely misspecify any generalized linear model. Since model misspecification can lead to biased coefficient estimates, this means that we could make incorrect inferences about the relationship between country-level factors and repressive states (Wooldridge 2001). Random Forests models, on the other hand, can account for non-linearities, other functional form possibilities, and interactions among variables (Breiman 2001).¹¹ As a consequence, these

¹⁰ Shmueli (2010) and Schrodt (2012) discuss this point in greater detail.

¹¹ Breiman (2001) describes these models in detail, and Jones and Linder (2016) discuss their many potential uses in political science research.

models are “often a much better option than an inflexible parametric model that is not fully implied by the theory” (Fariss and Jones 2015, 18).

Results and Discussion[NR2]

To determine whether the scalar assumptions are appropriate, or whether the results are more in line with one of the alternative expectations, “isolation”, “overlapping within levels”, or “overlapping across levels”, (or possibly a hybrid), we analyze how the dominant predictors of state repression (as identified in the existing literature), fare when we reconsider repression as a series of discrete outcomes using a series of Random Forests models.

Random Forest models have the disadvantage that the raw output from these models is difficult to understand, so we follow the standard practice (e.g., Jones and Linder 2016) of presenting the results of our classifiers graphically (see Figure 2). The far-left panel in Fig. 2 presents the results from a model with PTS 5 as the outcome measure, the far-right panel presents the results from a model with PTS 1 as the outcome, and the panels in between present the results from models that predict middle category levels. The vertical axis in each panel displays the variables we use in our models. The horizontal axis displays estimates of permutation accuracy for each variable. This is calculated as the difference in mean squared error between a model that is fitted using the observed values for a measure and a model that is fitted using random (but realistic) values for the same measure. This measure is then scaled to represent the percentage increase or decrease in mean square error caused by permuting the values of the variable. Theoretically, the measure spans both positive and negative values. The former indicate that variables increase the predictive performance of the model, while the latter indicate that variables decrease the predictive performance of the model. Higher positive values are associated with more important variables, while higher negative values are associated with less important variables. In line with Hill and

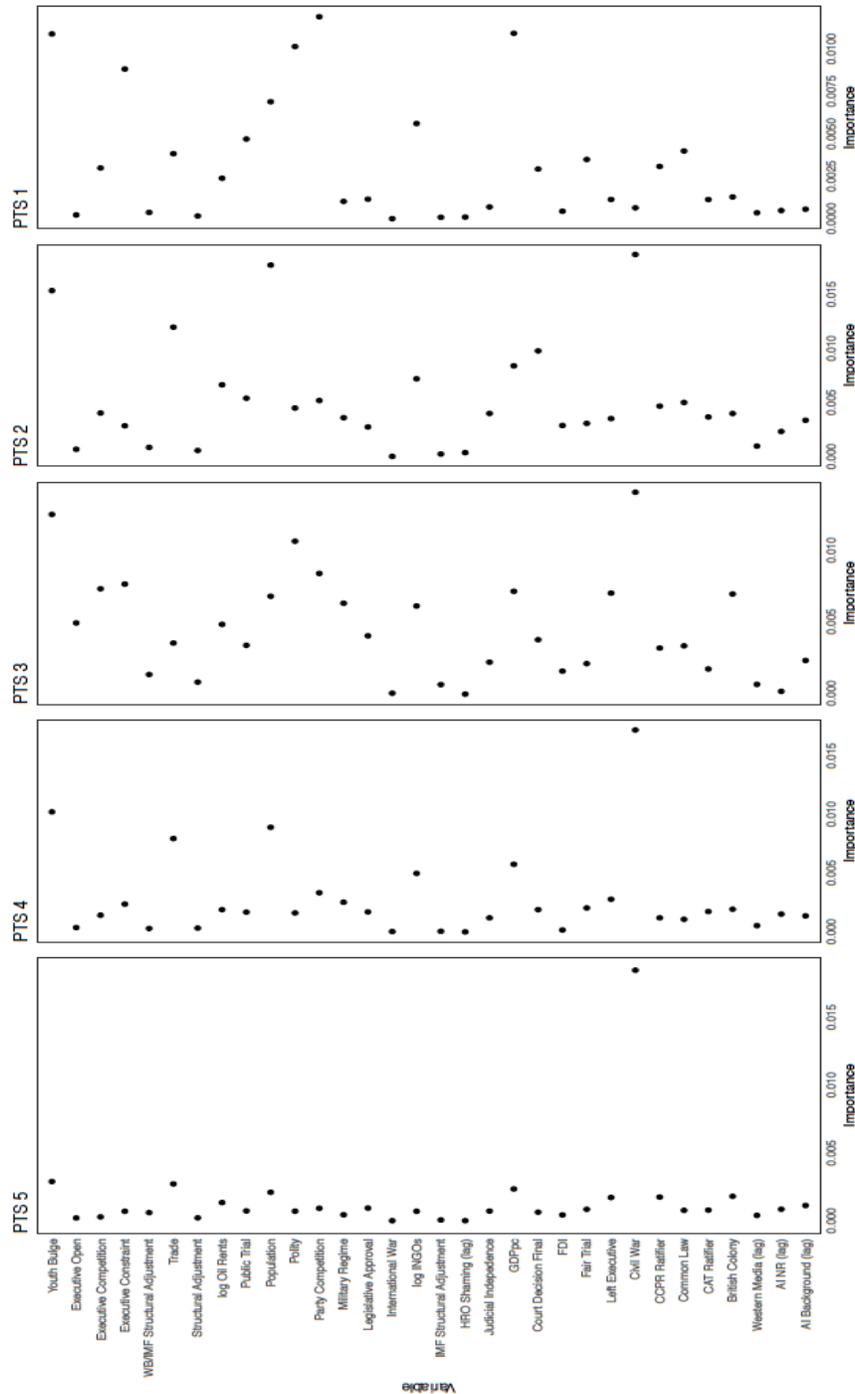
Jones (2016), Figure 4 (below) shows that all the variables that we include in each model improves its predictive importance. We interpret this as additional evidence that prior theoretical and empirical work has succeeded in identifying a set of factors that increase the accuracy (to some degree) of our predictions about levels of state repression.

Crucially, though, the relative importance of variables changes across our models. What this means is that some country-level characteristics are better predictors of certain levels of state respect for human rights than others.¹² Take, for example, *Civil War*, which is the most important predictor for PTS levels 5, 4, 3, and 2. On the other hand, knowing whether a country is engaged in domestic armed conflict, does little to improve our predictions about whether a country is scored a PTS 1. Similarly, we find that *Population* is the best predictor by far for *PTS 1* but is relatively less important for predicting other levels of state repression, and that the relative rank-order importance of *Youth Bulges* varies considerably across our dependent variables.¹³

¹² While readers might be interested in the effects of many variables in our models, we will focus our discussion of the results on larger patterns, mentioning individual variables only to the extent that they help support a broader point. When possible, we focus on those factors that have been found to be the most robust predictors of state repression – civil war, youth bulges, and population size (Hill and Jones 2014). One productive avenue for future research would be to further unpack the relationships we discover in our variable important plots.

¹³ In contrast to the finding that some variables are more important at predicting some categories of state repression than others, we also find that some variables are consistently not important predictors of PTS categories. While a growing literature would suggest that factors such as *Judicial Independence*, *International War*, and *Oil Rents* influence state repression, we find little evidence for that here. This is in line with the findings reported by Hill and Jones (2014) but builds on them by showing that these factors are not only poor predictors across repression levels but also poor predictors of individual categories. Taken together then, these sets of results suggest the need for a theoretical reappraisal of low-performing variables.

Figure 4. Predicting PTS Levels



Note: Figure 4 displays the results of five Random Forests classifiers. The outcome for each is a binary indicator that equals ‘1’ for a specific *PTS* outcome and ‘0’ otherwise. The vertical axis in each panel displays the variables in the model. The horizontal axis displays estimates of permutation accuracy for each variable.

At first glance, the finding that our predictors do not perform equally well across PTS categories might seem unsurprising or even unimportant. It has important consequences, though, for how we think about and conduct cross-national studies of state repression. One is that it highlights that some variables do more to explain some repression categories than others, a point largely not considered in the current literature. Indeed, our finding that variables have heterogeneous effects across outcomes suggests that the scalar assumption has partially blinded human rights scholars to interesting variation in the extent to which independent variables predict certain categories.

This finding also illustrates that our knowledge about the drivers of state repression is stronger for some categories than others. While it might initially look like these patterns are about the same, closer inspection of the plots reveals that this is not the case. That can be seen in the range of the horizontal axis in each of our plots. For two of them, this axis runs from 0 to approximately 0.01. As a reminder, this means that importance (i.e. permutation accuracy) of our independent variables ranges from a minimum around 0 to a maximum of 0.01. We see, though, that the importance of variables within the *PTS 2* plot range from 0 to about 0.02. This indicates that our variables do a much better job predicting if states are in this category of respect for human rights than if states are in others. Indeed, there are 3 variables in the *PTS 1* model that have greater predictive power than any variable in the other models (*Population, Trade, Youth Bulges*). In a broader view, these findings suggest that prior human rights work has led to the adoption of an empirical measuring and modeling approach that is better at predicting some categories or levels of government repression than others.

Additional support for this point comes from model fit statistics. Appendix B contains these statistics for the models presented in Figure 4 and the other models introduced below. We focus

here on each model's root-mean-square error (RMSE), a measure of predictive accuracy; lower values of RMSE indicate better model predictions. The RMSE for models with *PTS 5* and *PTS 1* as outcomes are 0.228 and 0.248, respectively; it is much higher, though, for other models, ranging from 0.330 for the *PTS 4* model to 0.408 for the *PTS 3* model. Again, this means that the empirical model we use is much better at predicting some states than others. In particular, it suggests that the full model is much better at predicting the best and worst categories but struggles with the middle outcomes.

More generally, the findings from Figure 4 shed new light on the determinants of human rights practices. Since the existing literature cannot easily account for these patterns, they suggest several new directions for theoretical development. But there is more to be gained from considering human rights states as discrete outcomes. Once we have adapted this view, we can start thinking about measuring repression in other, different ways.

Previously, the literature has examined changes across a scale, but we might also be interested in examining changes between different categories of state repression. We illustrate the utility of this approach below by examining what factors lead to a change in one PTS category to the next¹⁴ We focus on one-unit changes because these are the most commonly observed in the data; when a state's level of repression changes, it changes by one unit 69% of the time. We re-estimate a new set of Random Forests model with several new discrete (binary) indicators. These dummies are coded 1 if a country moves 'up' the PTS scale from one level to another.¹⁵ This re-

¹⁴ There are several reasons why scholars might want to do this instead of just examining predictors across a range of scale values. One reason is if scholars suspect that the levels are not ordered but unordered. This might be a reasonable assumption if researchers believe that the different levels capture movement along different dimensions.

¹⁵ In Appendix C, we provide a table of all category changes in our data, showing the frequency with which states move from one category of the PTS scale to another.

coding results in a set of four outcomes: *PTS 1 to PTS 2*, *PTS 2 to PTS 3*, *PTS 3 to PTS 4*, and *PTS 4 to PTS 5*.¹⁶ Figure 5 displays the results from this exploratory analysis, which reinforce some of the general findings from Figure 4.

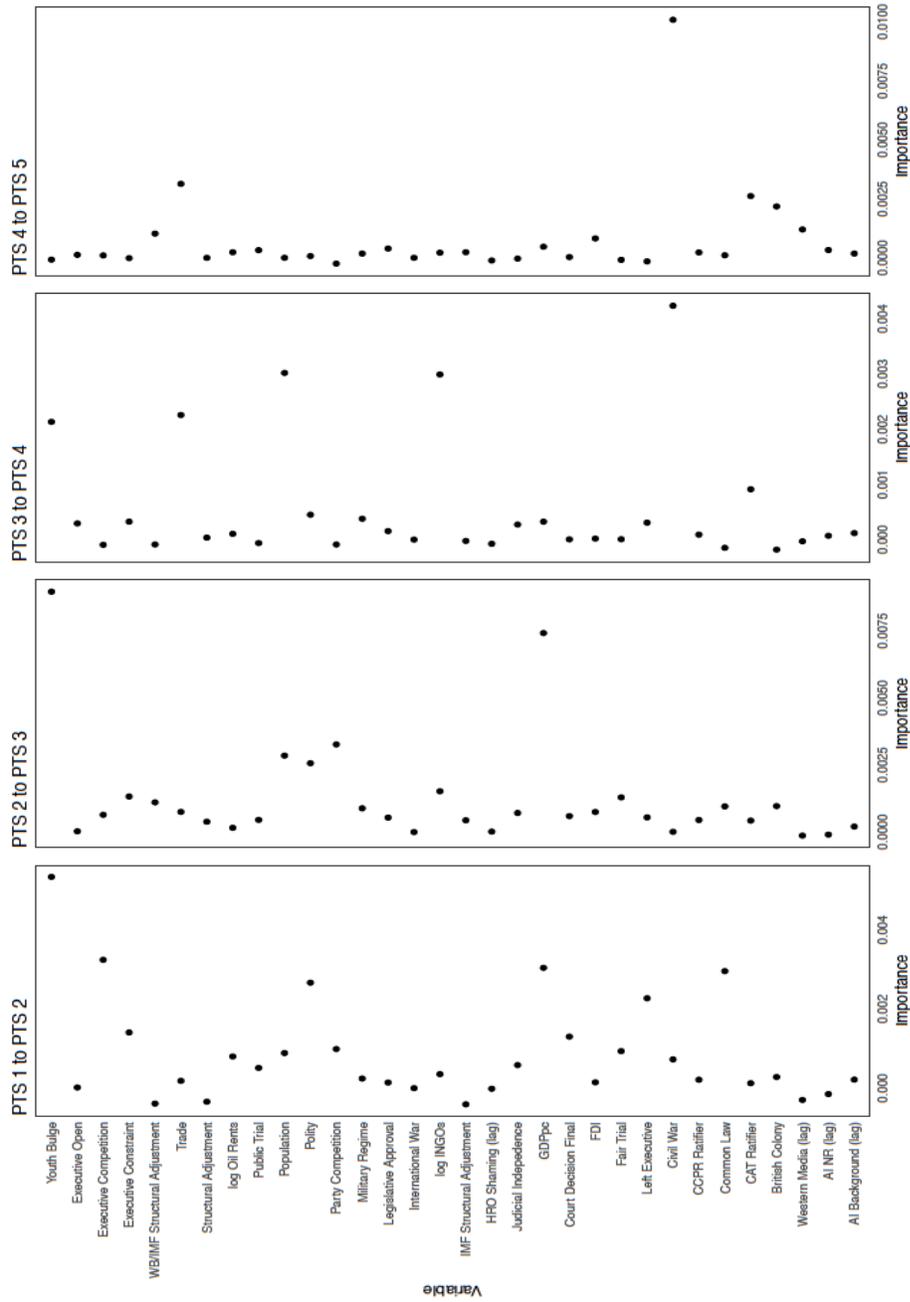
While Figure 4 shows that the importance of variables in predicting human rights practices varies considerably across discrete outcomes, Figure 5 indicates that some variables are important predictors for some PTS changes but not for others. The *Youth Bulge* measure, for instance, is the fifth most important predictor of a country moving from PTS 2 to PTS 1 but one of the least important predictors for a country moving from PTS 5 to PTS 4. Similarly, we see that the importance of *Population* fluctuates widely across outcomes. It is one of the most important factors for *PTS 5 to PTS 4* and *PTS 3 to PTS 2*, but relatively unimportant for *PTS 4 to PTS 3* and *PTS 2 to PTS 1*. These differential effects represent new findings that should help spur future theorizing and additional empirical work.

We also see, as we did in Figure 4, that the standard model of human rights abuse, taken as a whole, is better at predicting some of our transition outcomes than others. This is indicated by the range of values on the x-axis of these plots. For example, the range of Importance values is from but from only from 0 to 0.004 for *PTS 3 to PTS 4*. This indicates that we have only a partial understanding of what causes states to change their behavior between different repressive categories.¹⁷

¹⁶ One might also be interested in examining what predicts a country to move ‘down’ the PTS scale from a level to another. In other words, people might be interested in the predictors of these four outcomes: *PTS 5 to PTS 4*, *PTS 4 to PTS 3*, *PTS 3 to PTS 2*, and *PTS 2 - PTS 1*. With the hope of stimulating future research, we present the results of those analyses in Appendix D.

¹⁷ We also test the shifts from one level to the next in the more traditional way of a set of logit regressions, found in Appendix E. This shows how some variables are statistically significant predictors of some shifts but not others, and how some variables are not significant predictors of any shifts. These analyses also support the main proposition that the scalar assumption is problematic.

Figure 5. Predicting Category Changes



Note: Figure 5 displays the results of four Random Forests classifiers. The outcome for each is a binary indicator that equals ‘1’ when a country-year changes from a specific level on the PTS scale to another and ‘0’ otherwise. The vertical axis in each panel displays the variables in the model. The horizontal axis displays estimates of permutation accuracy for each variable. Unlike the prior analyses, the N for each model changes here. This is because we only examine whether countries coded at one category level change to another. The N for the models, moving from the left to the right, is 665, 665, 637, and 362.

Of the different conceptual categories in Figure 2, the results come closest to resembling the “Overlapping within levels” variety, and show some similarity to “overlapping across levels”. They are the farthest from the scalar approach (Figure 1) and “isolation”. This suggests both that repression (are conceived of in PTS) is not scalar (in the understanding presented in Figure 1), and that there are clear advantages to conceptualizing and modeling repression as discrete levels. However, the levels are not completely unrelated (corresponding to the “isolation” variety of thinking of discrete levels); but the relationships between levels and what drives repression is more complex than the scalar assumption has led on. More importantly, given these results, conclusions and policy implications drawn based on a scalar assumption are going to be highly problematic.

The results from these two groups of analyses (predicting particular outcomes or shifts between outcomes) highlight how much might be gained if we consider discrete human rights outcomes, instead of only ordered or interval ones. By extending our analysis of human rights outcomes in this way, we can examine both the predictors of (or, using more standard approaches, the correlates of) discrete levels/categories of state repression, but also imagine and analyze new measures, such as the transition from one human rights state to another. Our new way of conceptualizing and modeling repression as discrete outcomes leads to two main advantages over existing practices. First, this new modeling strategy allows us to more fully explore how different independent variables fare at predicting different repression outcomes. This is particularly useful for isolating what drives the outcomes we care about the most, namely instances of massive physical integrity rights abuses by states. Scholars following this measurement strategy will therefore be better suited to provide policy advice aimed at the specific goal of stopping mass atrocities.

Second, and perhaps more importantly, our approach to conceptualizing and modeling state repression as discrete outcomes suggests future studies should think more carefully about the theoretical mechanisms that link country-level factors to specific human rights abuses. Our hope is that this reinvigorates repression research, and helps researchers produce more sophisticated and/or insightful theoretical models than the relatively simple cost-benefit analysis and oversimplified threat-response assumption we find in most existing cross-national studies. Although we have learned a lot from the expansive literature on repression that has developed over the last few decades, the accumulation of knowledge in the cross-national study of repression has arguably slowed, in part because of stagnant theoretical development. Related, in the wake of a more sophisticated theoretical discussion in this field, we can also see the literature moving forward by reimagining the measures we include on the left-hand-side of our empirical models to better link empirics and theory.

Conclusion

Is state repression a difference in degree or difference in kind phenomenon? And, what are the larger implications of asking that question and how we answer it? In this paper, we show the repression literature has conceptualized state repression as a difference in degree phenomenon, what we refer to as the scalar assumption. At the same time, and largely as a consequence of the former, the repression literature has a blind spot to the question of whether the same factors indeed drive different levels of repression, having historically examined what drives states physical integrity violations under the assumption that the answer to that question is yes. The heretofore unchallenged assumption that repression is a scalar phenomenon, with equidistance between the categories and the same drivers affecting all stages of the scale equally, has dominated large- N

studies of state repression. It has also structured how we think about and give policy solutions to counter repression. We challenge this scalar paradigm. Specifically, we relax that assumption and provide an alternative way of seeing repression as discrete outcomes. The point we are making in this paper is part of the general critical discussions about how we model state repression. More fundamentally, however, our critique focus on how the ways in which repression is conceptualized (and therefore modeled) is based on potentially problematic assumptions which could lead to faulty inferences and conclusions about findings, and how a new conceptualization will force us to think anew about theories of state repression.

We demonstrate that different levels of repression are not driven by the same factors, and present findings are at odds with the scalar assumption that threads through existing repression scholarship. We can therefore conclude that our findings highlight the need for greater theorizing about the relationships between our independent variables and different categories or levels of repression, and that rethinking repression as discrete outcomes is a key factor in that pursuit. At the moment, the existing literature offers little explanation for these more specific patterns. Another key implication from this reorientation and our analysis is that, existing cross-national studies are not yet able to provide robust suggestions for how to counter large-scale state repression, including genocides and mass atrocities. Key theoretical variables are unable to predict the highest levels of human rights abuses (PTS 5), and *no* independent variable can consistently predict all levels of state repression. Our theoretical discussion and findings therefore provide strong justification for discretizing existing human rights scales in future work. We hope the approach of conceiving of and modeling repression as a set of discrete levels will open new lines of inquiry, and prompt researchers to think more carefully about the theoretical mechanisms that link country level factors to specific levels of human rights abuses.

References

- Abouharb, M. Rodwan., and Cingranelli, David. 2007. *Human rights and structural adjustment*. Cambridge University Press.
- Apodaca, Clair. 2001. "Global Economic Patterns and Personal Integrity Rights after the Cold War." *International Studies Quarterly* 45 (4): 587–602.
- Bagozzi, Benjamin E., and Daniel Berliner. "The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of US State Department Human Rights Reports." *Political Science Research and Methods* (2016): 1-17.
- Barsh, Russel Lawrence. 1993. "Measuring human rights: Problems of methodology and purpose." *Hum. Rts. Q.* 15: 87.
- Beck, Thorsten. 2000. *New tools and new tests in comparative political economy: The Database of Political Institutions*. World Bank Publications.
- Bollen, Kenneth A. "Issues in the comparative measurement of political democracy." *American Sociological Review* (1980): 370-390.
- Bollen, Kenneth A. "Political rights and political liberties in nations: An evaluation of human rights measures, 1950 to 1984." *Hum. Rts. Q.* 8 (1986): 567.
- Bollen, Kenneth A. "Political democracy: Conceptual and measurement traps." *Studies in Comparative International Development (SCID)* 25.1 (1990): 7-24.
- Bollen, Kenneth. "Liberal democracy: Validity and method factors in cross-national measures." *American Journal of Political Science* (1993): 1207-1230.
- Booyesen, Frederik. "An overview and evaluation of composite indices of development." *Social indicators research* 59.2 (2002): 115-151.

- Breiman, Leo. 2001. "Random forests." *Machine Learning*, 45(1), 5-32.
- Bueno de Mesquita, Bruce, George W. Downs, Alastair Smith, and Feryal Marie Cherif. 2005. "Thinking Inside the Box: A Closer Look at Democracy and Human Rights." *International Studies Quarterly* 49: 439–57.
- Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2010. "Democracy and dictatorship revisited." *Public Choice* 143(1-2): 67-101.
- Cingranelli, David L. and David L. Richards, 1999. *Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights*, 43 INT'L STUD. Q. 407.
- Cingranelli, David L., and David L. Richards. 2010. "The Cingranelli and Richards (CIRI) human rights data project." *Human Rights Quarterly* 32(2): 401-424.
- Collier, David, and Robert Adcock. "Democracy and dichotomies: A pragmatic approach to choices about concepts." *Annual Review of Political Science* 2.1 (1999): 537-565.
- Coppedge, Michael, et al. 2016. "Varieties of Democracy: Codebook v6." Varieties of Democracy (V-Dem) Project. URL: <https://www.v-dem.net/en/reference/version-6-mar-2016/>
- Conrad, Courtenay Ryals, and Will H. Moore. "What stops the torture?." *American Journal of Political Science* 54.2 (2010): 459-476.
- Conrad, Courtenay R., Jillienne Haglund, and Will H. Moore. 2013. "Disaggregating torture allegations: Introducing the ill-treatment and torture (ITT) country-year data." *International Studies Perspectives* 14(2): 199-220.
- Coppedge, Michael., Gerring, John., Lindberg, Steven. 1990. Varieties of Democracy. *Studies in Comparative International Development*, 25(1), 51-72.
- Dallin, Alexander, and George W. Breslauer. *Political terror in communist systems*. Stanford: Stanford University Press, 1970.

- Davenport, Christian. "The weight of the past: Exploring lagged determinants of political repression." *Political Research Quarterly* 49.2 (1996): 377-403.
- Davenport, Christian. 1997. "From Ballots to Bullets: An Empirical Assessment of How National Elections Influence State Uses of Political Repression." *Electoral Studies* 16 (4): 517-40.
- Davenport, Christian. 2007. "State repression and political order." *Annual Review Political Science*, 10: 1-23.
- Davenport, Christian. 2009. *Media bias, perspective, and state repression: The Black Panther Party*. Cambridge University Press.
- Davenport, Christian, Sarah Soule and David Armstrong. 2011. "Protesting While Black? The Differential Policing of American Activism, 1960-1990." *American Sociological Review* 76(1): 152-178.
- De Soysa, Indra, and Ragnhild Nordås. "Islam's bloody innards? Religion and political terror, 1980-2000." *International Studies Quarterly* 51.4 (2007): 927-943.
- Earl, Jennifer. "Political repression: Iron fists, velvet gloves, and diffuse control." *Annual review of sociology* 37 (2011): 261-284.
- Fariss, Christopher J., and Zachary M. Jones. 2017. "Enhancing validity in observational settings when replication is not possible." *Political Science Research and Methods*: 1-16.
- Fariss, C.J., 2014. Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, 108(2), pp.297-318.
- Fein, Helen. 1995. "More Murder in the Middle: Life-Integrity Violations and Democracy in the World, 1987." *Human Rights Quarterly* 17 (1): 170-92.
- Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52 (3): 647-74.

- Gleditsch, Nils. P., Wallensteen, Peter., Eriksson, Mikael., Sollenberg, Margareta, and Strand, Håvard. 2002. "Armed conflict 1946-2001: A new dataset." *Journal of Peace Research*, 39(5), 615-637.
- Hafner-Burton, Emilie M. "Trading human rights: How preferential trade agreements influence government repression." *International Organization* 59.3 (2005): 593-629.
- Hafner-Burton, Emilie M. "Right or robust? The sensitive nature of repression to globalization." *Journal of Peace Research* 42.6 (2005): 679-698.
- Hafner-Burton, Emilie. 2008. "Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem." *International Organization* 62: 689-716.
- Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. "Human rights in a globalizing world: The paradox of empty promises." *American Journal of Sociology* 110(5): 1373-1411.
- Harff, Barbara. "No lessons learned from the Holocaust? Assessing risks of genocide and political mass murder since 1955." *American Political Science Review* 97.1 (2003): 57-73.
- Hathaway, Oona A. "Do human rights treaties make a difference?." *The Yale Law Journal* 111.8 (2002): 1935-2042.
- Henderson, Conway W. "Conditions affecting the use of political repression." *Journal of Conflict Resolution* 35.1 (1991): 120-142.
- Hibbs Jr, Douglas A. 1973. "Problems of Statistical Estimation and Causal Inference in Time-Series Regression Models." *Sociological Methodology* 1974: 252-308.
- Hill, Danny. W., and Jones, Zachary. M. 2014. "An empirical evaluation of explanations for state repression." *American Political Science Review*, 108(3), 661-687.
- Hofstede, Geert. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications, 2003.

- Jones, Zachary M. 2014. "untreaties: scrapes, cleans, transforms, and searches United Nations treaty data." <https://github.com/zmjones/untreaties>
- Jones, Zachary M. and Fridolin Linder. 2016. "edarf: Exploratory Data Analysis using Random Forests." *Journal of Open Source Software*.
- Kalyvas, Stathis N. *The logic of violence in civil war*. Cambridge University Press, 2006.
- Keith, Linda Camp. "The United Nations International Covenant on Civil and Political Rights: Does it make a difference in human rights behavior?." *Journal of Peace Research* 36.1 (1999): 95-118.
- Keith, Linda Camp, C. Neal Tate, and Steven C. Poe. 2009. "Is the law a mere parchment barrier to human rights abuse?" *The Journal of Politics*,71(2): 644-660.
- Knack, Stephen, and Philip Keefer. "Institutions and economic performance: cross-country tests using alternative institutional measures." *Economics and Politics* 7.3 (1995): 207-227.
- Krain, Matthew. "State-sponsored mass murder: The onset and severity of genocides and politicides." *Journal of conflict resolution* 41.3 (1997): 331-360.
- Landman, Todd. 2004. "Measuring human rights: principle, practice and policy." *Human Rights Quarterly* 26.4: 906-931.
- Landman, Todd, and Edzia Carvalho. 2009. *Measuring human rights*. Routledge.
- Linz, Juan J. 1985. *Totalitarian and authoritarian regimes*. Addison-Wesley.
- Marshall, Monty G., and Keith Jagers. 2002. "Polity IV project: Political regime characteristics and transitions, 1800-2002."
- McCormick, James M., and Neil J. Mitchell. 1997. "Human rights violations, umbrella concepts, and empirical analysis." *World Politics* 49(4): 510-525.

- Mitchell, Neil J., and James M. McCormick. "Economic and political explanations of human rights violations." *World Politics: A Quarterly Journal of International Relations* (1988): 476-498.
- Mitchell, Sara McLaughlin, Jonathan J. Ring, and Mary K. Spellman. 2013. "Domestic legal traditions and states' human rights practices." *Journal of Peace Research* 50(2): 189-202.
- Murdie, Amanda M., and David R. Davis. 2012. "Shaming and blaming: Using events data to assess the impact of human rights INGOs." *International Studies Quarterly* 56(1): 1-16.
- Neumayer, Eric. 2005. "Do international human rights treaties improve respect for human rights?." *Journal of Conflict Resolution* 49(6): 925-953.
- Nordås, Ragnhild, and Christian Davenport. 2013. "Fight the youth: Youth bulges and state repression." *American Journal of Political Science* 57(4): 926-940.
- Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. "Democratic compromise: A latent variable analysis of ten measures of regime type." *Political Analysis* 18(4): 426-449.
- Poe, Steven C., and C. Neal Tate. "Repression of human rights to personal integrity in the 1980s: A global analysis." *American Political Science Review* 88.4 (1994): 853-872.
- Poe, Steven C., C. Neal Tate, and Linda Camp Keith. 1999. "Repression of the Human Right to Personal Integrity Revisited: A Global Cross-National Study Covering the Years 1976–1993." *International Studies Quarterly* 43(2): 291-313.
- Poe, Steven C., Sabine C. Carey, and Tanya C. Vazquez. 2001. "How are these pictures different? A quantitative comparison of the US State Department and Amnesty International human rights reports, 1976-1995." *Human Rights Quarterly* 23(3): 650-677.
- Richards, David. 1999. "Perilous Proxy: Human Rights and the Presence of National Elections." *Social Science Quarterly* 80(4): 648-665.

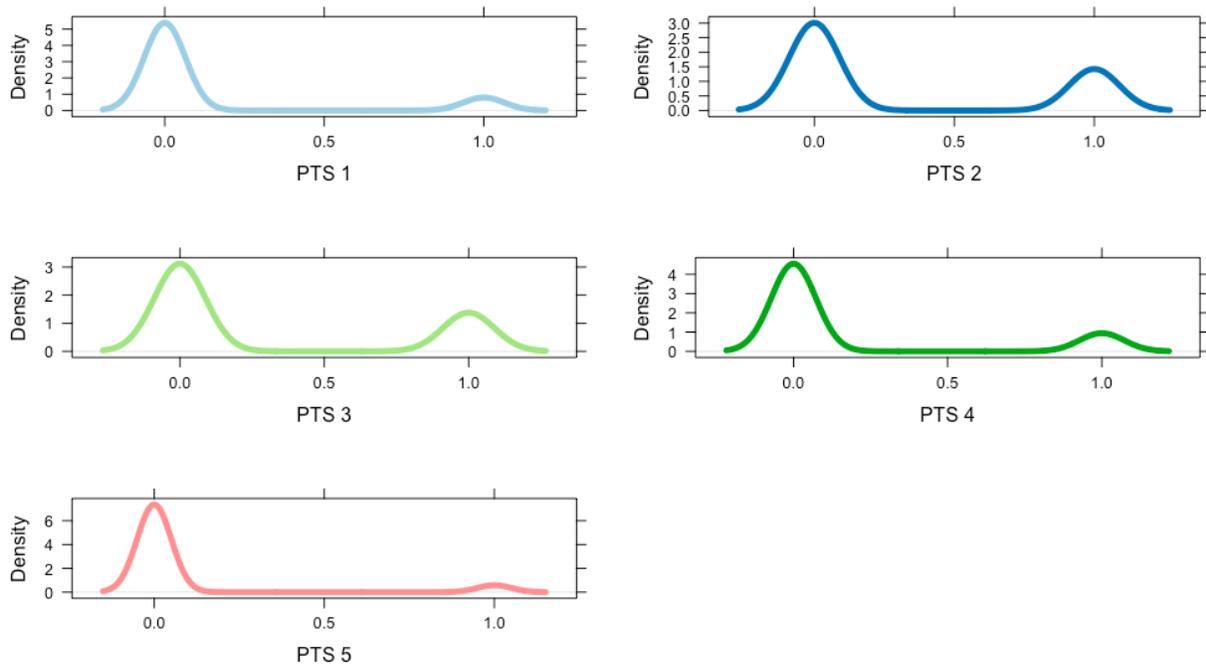
- Richards, David L., Ronald D. Gelleny, and David H. Sacko. "Money with a mean streak? Foreign economic penetration and government respect for human rights in developing countries." *International Studies Quarterly* 45.2 (2001): 219-239.
- Ritter, Emily Hencken, and Courtenay R. Conrad. "Preventing and responding to dissent: The observational challenges of explaining strategic repression." *American Political Science Review* 110.1 (2016): 85-99.
- Ron, James, Howard Ramos, and Kathleen Rodgers. 2005. "Transnational information politics: NGO human rights reporting, 1986–2000." *International Studies Quarterly* 49(3): 557-587.
- Ross, Michael. 2006. "A closer look at oil, diamonds, and civil war." *Annual Review of Political Science*, 9: 265-300.
- Schrodt, Philip A. 2014. "Seven deadly sins of contemporary quantitative political analysis." *Journal of Peace Research* 51(2): 287-300.
- Shmueli, Galit. 2010. "To explain or to predict?." *Statistical Science*, 25(3): 289-310.
- Spilker, Gabriele, and Tobias Böhmelt. 2013. "The impact of preferential trade agreements on governmental repression revisited." *The Review of International Organizations* 8(3): 343-361.
- Stohl, Michael, and George A. Lopez, eds. *Government violence and repression: an agenda for research*. No. 148. New York: Greenwood Press, 1986.
- Stohl, Michael, et al. "State violation of human rights: Issues and problems of measurement." *Hum. Rts. Q.* 8 (1986): 592.

- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. "Conditional variable importance for random forests." *BMC bioinformatics*, 9(1), 307.
- Tilly, Charles. "Event catalogs as theories." *Sociological Theory* 20.2 (2002): 248-254.
- Urdal, Henrik. 2006. "A clash of generations? Youth bulges and political violence." *International Studies Quarterly*, 50(3): 607-629.
- Valentino, Benjamin, Paul Huth, and Dylan Balch-Lindsay. "'Draining the sea': mass killing and guerrilla warfare." *International Organization* 58.2 (2004): 375-407.
- Vanhanen, Tatu. "Domestic ethnic conflict and ethnic nepotism: A comparative analysis." *Journal of Peace Research* 36.1 (1999): 55-73.
- Verdeja, Ernesto. "The political science of genocide: outlines of an emerging research agenda." *Perspectives on Politics* 10.2 (2012): 307-321.
- Walter E.V. 1969. *Terror and Resistance; a Study of Political Violence, with Case Studies of Some Primitive African Communities*. New York: Oxford Univ. Press
- Wood, Reed M. 2008. "'A Hand upon the Throat of the Nation': Economic sanctions and State Repression, 1976-2001." *International Studies Quarterly* 52: 489-513.
- Wood, Reed. M., and Gibney, Mark. 2010. "The Political Terror Scale (PTS): A re-introduction and a comparison to CIRI." *Human Rights Quarterly*, 32(2), 367-400.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- Zanger, Sabine C. "A global analysis of the effect of political regime changes on life integrity violations, 1977-93." *Journal of Peace Research* 37.2 (2000): 213-233.

Supplementary Appendices

Appendix A

Figure A1: Density Plots of Outcome Measures



Note: Figure A1 presents density plots for each of our outcome measures. These are constructed from our dataset of 2,224 observations, covering 157 countries from 1981-1999.

Appendix B

Table A1: Random Forests Model Fit Statistics

	RMSE	R²	MAE
PTS 5	0.228	0.284	0.081
PTS 4	0.330	0.248	0.202
PTS 3	0.408	0.234	0.339
PTS 2	0.403	0.268	0.335
PTS 1	0.248	0.464	0.109
PTS 1 to 2	0.466	0.022	0.429
PTS 2 to 3	0.380	0.075	0.275
PTS 3 to 4	0.339	0.056	0.201
PTS 4 to 5	0.344	0.078	0.213
PTS 5 to 4	0.490	0.003	0.470
PTS 4 to 3	0.434	0.042	0.367
PTS 3 to 2	0.404	0.076	0.314
PTS 2 to 1	0.293	0.087	0.146

Note: Table A1 presents several fit statistics for the Random Forests models estimated in the main text. The columns denote three different fit statistics. The rows denote outcomes.

Appendix C

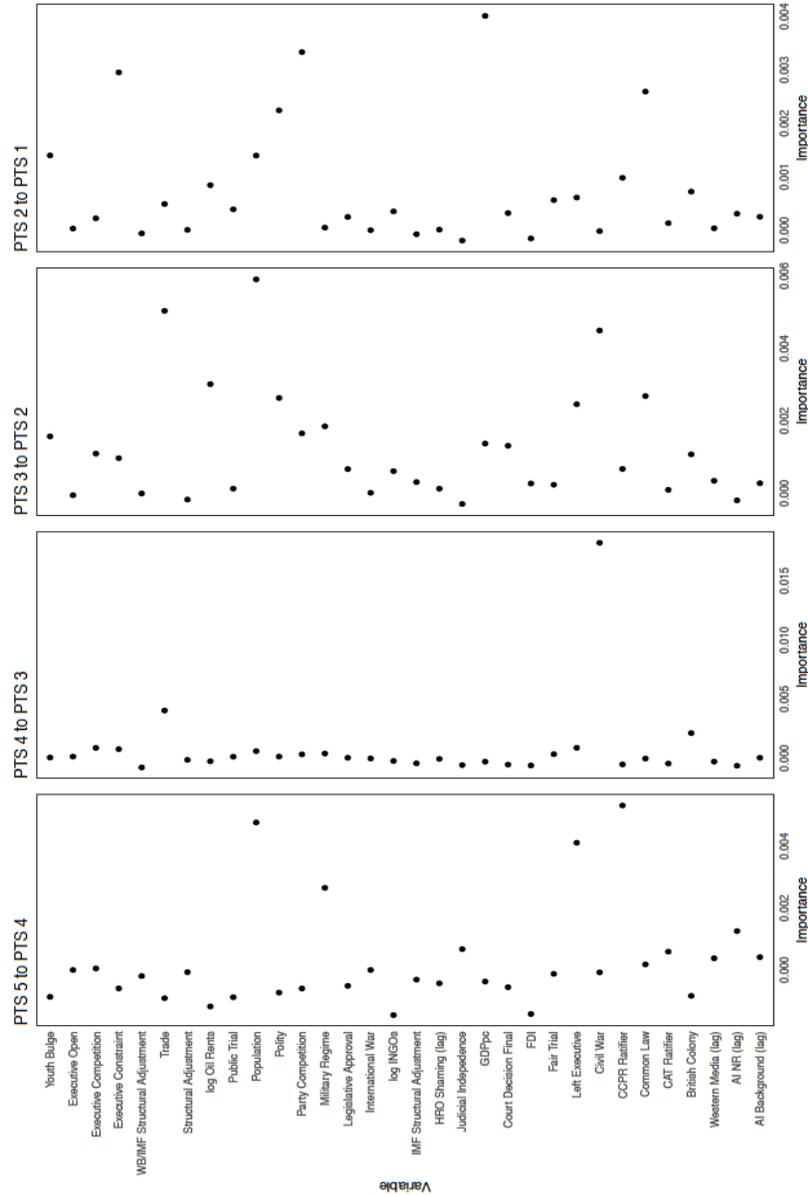
Table A2: Year-by-Year Category Changes

	PTS 5 (<i>t-1</i>)	PTS 4 (<i>t-1</i>)	PTS 3 (<i>t-1</i>)	PTS 2 (<i>t-1</i>)	PTS 1 (<i>t-1</i>)
PTS 5 (<i>t</i>)	95	65	23	13	11
PTS 4 (<i>t</i>)	73	212	104	37	15
PTS 3 (<i>t</i>)	57	151	441	182	65
PTS 2 (<i>t</i>)	47	57	189	488	109
PTS 1 (<i>t</i>)	33	34	37	102	212

Note: Table A2 shows the frequency with which states move from one category in the PTS scale to another from year-to-year. Rows denote country-year categories at time *t*, while columns denote country-year categories at time *t-1*. Highlighted cells indicate country-years in which state repressive practices do not change.

Appendix D

Figure A2: Predicting Shifts Using Random Forest Classifiers



Note: Figure A2 displays the results of four Random Forests classifiers. The outcome for each is a binary indicator that equals ‘1’ when a country-year changes from a specific level on the PTS scale to another and ‘0’ otherwise. The vertical axis in each panel ranks variables from top to bottom in order of their importance for predicting whether states exhibit a specific level of repression. The horizontal axis displays estimates of permutation accuracy for each variable. Unlike the prior analyses, the N for each model changes here. This is because we only examine whether countries coded at one category level change to another. The N for the models, moving from the left to the right, is 148, 362, 637, and 665.

Appendix E

Figure A3: Predicting Shifts Using Logit Regression (PTS based on Amnesty International)

	1	2	3	4
	(1-2)	(2-3)	(3-4)	(4-5)
DV: PTS(a)				
Polity IV= 8/9	-0.352** (0.158)	-0.277** (0.134)	-0.045 (0.126)	-0.258 (0.220)
Polity IV=10	-1.059*** (0.174)	-1.194*** (0.214)	-0.363 (0.249)	-0.538 (0.691)
Dissent	0.035 (0.024)	0.023 (0.015)	0.012 (0.011)	0.001 (0.012)
Civil conflict	0.172 (0.229)	0.819*** (0.140)	0.867*** (0.103)	0.660*** (0.110)
Population, ln	0.090* (0.053)	0.128*** (0.038)	0.070** (0.034)	-0.025 (0.041)
Youth bulge	0.010 (0.015)	0.038*** (0.012)	0.028*** (0.010)	-0.003 (0.014)
GDP/capita, ln	-0.127 (0.083)	0.127 (0.084)	0.031 (0.068)	0.029 (0.070)
LDV	1.183*** (0.147)	1.087*** (0.077)	1.097*** (0.063)	0.968*** (0.075)
N	2,586	2,586	2,586	2,586

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1