

American Political Science Review  
December 7, 2013

Dear Editors and Reviewers,

We would like to thank you all for the opportunity to revise our manuscript, “An Empirical Evaluation of Explanations for State Repression” (Ms. No. APSR-D-13-00597). We have made many revisions to the paper along the lines suggested by the editors and reviewers and we believe the manuscript has improved significantly as a result.

We have worked especially hard to broaden the appeal of the manuscript beyond those interested primarily in state repression, a point emphasized by both the editors and Reviewer 2. Our efforts in this regard include, first, explaining in the introduction and literature review how empirical work on state repression is connected to other literatures in political science, particularly those on democratization and institutional constraints on government predation. The incorporation of this literature into our argument also helps give the paper a clearer theoretical framework through which our results can be interpreted, which addresses Reviewer 1’s primary concern. Second, we have further elaborated our point about statistical significance vs. prediction, and importantly point out that this problem characterizes other areas of political science research. We believe that non-human rights scholars will now find the paper more interesting. We have added a paragraph to the conclusion discussing how researchers in all subfields can address this problem in their own work.

Related, we have tried to make the discussion of the methods used in the paper (particularly the random forests) more accessible and intuitive. Additionally, we discuss in the conclusion the basic steps researchers in all subfields can take to incorporate some assessment of predictive power into their research designs. We believe this will also broaden the paper’s appeal.

We have also expanded our point about the conceptualization/measurement problems with using Polity in models of state repression, and present this argument in the introduction and literature review. This is in response to comments from the editors and Reviewer 1, and we believe this addition also gives the paper a better theoretical frame to help readers sort through the (many) results we present.

Below we discuss in more detail these changes and the many other changes we have made to the manuscript in response to comments from the editors and reviewers. We thank the editors again for the opportunity to revise the manuscript, and the reviewers for their extremely helpful comments.

Sincerely, The Authors

## Editor Comments and Responses

1) The editor writes: Provide a better introduction and theoretical framing. Explain why state repression is an important subject and why it is important to re-evaluate the significance of the “conventional” explanations for state repression. Keep in mind that the Review’s readership includes a broad cross-section of political scientists and not only specialists in the field of human rights. Explain to this readership why state repression matters and why it matters how we evaluate the explanatory value of different variables.

### **Response:**

We have made several changes to the manuscript that we think will help broaden the paper’s appeal to non-human rights scholars as well as better connect our results to existing theoretical arguments. We now discuss in the introduction (pp. 1, 4) and Section 2 (pp. 7-8) the connection between this literature and the literature on the emergence of democracy and institutional constraints on government more broadly. We have added a brief discussion of the comparative institutions literature on constitutions and courts to Section 2, and explain how this literature generates theoretical expectations regarding the relationship between repression and legal institutions, particularly constitutional provisions and independent judiciaries. Accordingly, we also note (pp. 27, 35) that the results we obtain for legal systems are largely in line with these expectations, and we discuss in the conclusion (pp. 35-36) the implications of these findings for future theoretical work on repression, in particular the likely usefulness of theories of judicial independence/power, and the connection between common law systems and judicial independence. We have also expanded the discussion of the methods we use to broaden the appeal of the manuscript. As we elaborate below, we point out that the research design problems we discuss and address are common outside of the literature on state repression.

2) The editor writes: Reviewer 3 would like to see the discussion of random forests to be made more accessible to the non-specialist.

### **Response:**

We significantly expanded the description of decision trees and random forests with the aim of making the method more accessible to the non-specialist. We now include a non-technical but detailed description of the algorithm. We now include a detailed empirical example on human rights data, which includes an informative figure, which we think compliments the verbal description. We have

also added further non-technical information about how random forests generally provide some advantages over the models we use for cross-validation. We also direct readers towards the introduction to recursive partitioning methods written by the authors of the particular method we use.

3) The editor writes: Quite late in the paper (on page 33) you argue that statistical significance is not the same as predictive power (and you reference literature on this point). It seems to us that this should be highlighted better as it pertains to the core of the paper's argument: although a number of variables are conventionally included in models explaining state repression and are consistently statistically significant, the variables may not hold much predictive value regarding when governments resort to repression.

**Response:**

We agree that this argument is central to the paper, and we now discuss this problem in the introduction (p. 2), and cite relevant conflict literature on this point. We have also significantly expanded the beginning of Section 3 (pp. 9-11). We explain in more detail why this is so important for the literature on state repression, and argue that a (nearly) complete lack of attention to model fit has caused researchers to draw erroneous conclusions about the accuracy of theoretical explanations for government violence. We also point out that, since researchers are building potentially inaccurate models, much of this work is not useful for informing policy. We also explain more thoroughly the issue of overfitting, and why evaluating *out-of-sample* fit is crucial to address this problem. We make the case that our analysis offers a better way to adjudicate between different theoretical explanations than conducting null hypothesis tests. Finally, we suggest that the methods we use could be usefully applied to any area of political science and offer practical advice about relatively easy ways to assess model performance other than significance tests.

4) The editor writes: Reviewer 1 suggests that you might discuss the measurement issues with the Polity score upfront, rather than wait until the discussion and conclusion. We encourage you to incorporate this into your argument.

**Response:**

We agree that this is an important point that we should emphasize early and often, and we now address this problem at several points throughout the paper. We believe this will help make apparent the implications of our results for theoretical arguments about repression, and also provides an anchor for readers to use in

interpreting some of our results. We first discuss the problems created by using the Polity scale in studies of repression in the introduction (pp. 4). We have also added a paragraph to Section 2 (pp. 5-6) which notes the conceptual connection between Polity and Dahl's (1971) conception of democracy, and explains why Polity is tautologically related to some existing measures of repression. We discuss this problem again in our discussion of the data we use (pp. 16, 20), and point out that this is an enormous problem for the "participation-competition" component of Polity and the CIRI scale that measures political imprisonment. We return to this point in the conclusion (pp. 33-35), and offer some advice to researchers about how to avoid this problem in future work.

5) The editor writes: There are a large number of figures in the paper. We understand their function, but would prefer that the paper is not cluttered with too many large-scale figures. We suggest two possible alternatives: 1. Keep all the figures, but present figures 3-16 together as a composite set of figures on one or two pages (which would also make it easier to spot the similarities alluded to in the paper's argument), and the same with figure 17-22; 2. Include only a few key figures in the text and present the complete sets in an (online) appendix.

**Response:**

For each set of model specifications (e.g. linear models with population and gdp, linear models with population, gdp and civil war, ordinal logistic regression with population and gdp, etc) we have created a combined plot that shows the increase in predictive power across all of the relevant measures of repression. This dramatically reduces the number of figures and makes it easier to compare the performance of independent variables across different measures of repression, which further highlights some of our theoretical insights.

6) The editor writes: Although a small point, we would like you to modify a sentence on page 15 to eliminate the phrase "we feel" and simply state why the rule you employ is justified.

**Response:**

We have changed the language used in this sentence as suggested, and simply note that the decision rule is justified given the sparse number of covariates included in each of the baseline models.

## **Reviewer 1 Comments and Responses**

1) Reviewer 1 writes: I do not think theory should be ignored. For example, I think that the

article should be framed around some theoretical approach: e.g., rational choice, principal-agent arguments or prospect theory - using these to help develop theoretical expectations which would in turn help sort results. At present, it is not clear what the author thinks about why repression is applied and how the various independent variables influence this expectation. Theory would help the author juxtapose domestic vs. international factors as well as diverse aspects of political institutions. These differences are found in the empirical analysis but in the current version of the manuscript it is not quite clear what one is to do with the diversity. At present, these divisions are ignored and the author acts as if all variables are theoretically equivalent (i.e., that they are all equally likely to be supported). Even the studies conducted are not of equal value. For example, I would argue that all of the internationally-oriented articles mis-specify the domestic factors highlighted by Bueno De Mesquita et al. as well as Davenport and Armstrong.

Theory would help the author think about state repression. For instance, at present there is no way to think about the various repressive indicators used. The CIRI subcomponents are perhaps reflective of important dimensions: overt vs. covert, constraining vs. eliminating citizens/challengers and selective vs. indiscriminate behavior. PTS combines scope, lethality and the degree of political targeting. It is possible that the different CIRI measures provide insight into what is driving PTS. I do not like the fact that the different measures are simply tossed in without thinking about what they are intended to measure and why different variables influence/don't influence them.

**Response:**

It is not our intention to offer a new theory of state repression, but we think this is a valuable suggestion, as the front-end of the previous manuscript gave little guidance to the reader in terms of interpreting results in light of existing theoretical expectations. As we mention above, we have now connected our results to the comparative institutions literature, which serves as a nice theoretical frame through which to interpret the results with respect to legal systems. These are among the strongest and most interesting results we present, and we think the manuscript benefits from this additional theoretical framing.

We have also added some discussion to the literature review (p. 6) to contrast international and domestic factors. We note that domestic factors generally receive far stronger empirical support in the previous literature, and that this difference shows up in our analysis. We reiterate this point in the results (pp. 28-29). Additionally, we briefly mention (p. 6) the theoretical motivations for some of the more IPE-related studies. Further, we discuss (p. 27) the overt/covert dimension of government violence raised by Reviewer 1 when discussing the marked

differences between our results for the CIRI scales measuring political imprisonment/torture and those measuring disappearances/killings. And, as discussed above, we start developing our point about the Polity scale in the first section of the manuscript. This creates some expectations regarding the competition aspect of democracy and the CIRI political imprisonment scale, which we outline on page 16.

2) Reviewer 1 writes: I do not find keeping lagged repression out of the model compelling. The theoretical explanation being drawn upon here includes bureaucratic inertia. Whether or not diverse variables can add explanatory power when the past is considered seems important. Historically, this is one of the variables that has mattered the most.

**Response:**

Upon reflection, we agree that this past result is more strongly tied to theory than the previous manuscript indicated. We have added a third baseline model to our cross-validation analysis, and discuss this decision on pages 18-19. We also cite relevant theoretical arguments which explain the importance of past repression for the use of repression in the present. We have necessarily removed the paragraph where we discussed our decision to exclude a lagged dependent variable from the analysis. A discussion of the results from this new analysis appears on page 28. We have also created figures to display the results from the cross-validation analysis using lagged repression, but we have chosen to include these in the appendix for two reasons. One is that the results are very consistent with the results from our other baseline models, which we note in the text. The other reason is that the inclusion of the lag soaks up much of the variance in the dependent variable, thus even fewer covariates add predictive power. The covariates that do add power are a subset of those that add power to the other baseline specifications, and we note in the text which of these still perform well using the third baseline specification.

3) Reviewer 1 writes: I think that the measurement issues need to be addressed. I immediately thought that the use of Polity was problematic because I knew that the sub-dimensions include elements of repression. Why wait for putting this in the conclusion. Related to this though is the problem of civil war. I realize that previous work has included this in their models but given that repressive behavior is part of the concept, this seems conceptually unacceptable. Run models without this variable in addition to noting this conceptual problem.

**Response:**

As we discuss above, we now discuss the issues with Polity in the introduction and at several subsequent points in the manuscript. We think Reviewer 1 makes an excellent point that similar conceptualization and measurement problems affect civil war. We now mention these problems in the introduction (pp. 3-4) and results (p. 24) sections and discuss them more thoroughly in the conclusion (pp. 32-33). We point out that civil war deaths inevitably include non-combatants and so overlap, conceptually and empirically, with measures of state repression, at least those that measure lethal uses of force. In the conclusion (p. 33) we go so far as to encourage researchers to exclude measures of civil war from statistical models of state repression.

4) Reviewer 1 writes: I'm wondering if the different models employed can be tied to specific theoretical explanations. Do the different approaches provides insights into the use of state repression through their assumptions, ability to consider interactions or some other criteria?

**Response:**

As discussed above, we have added substantially to the paper's theoretical frame, and hope this helps the reader sort through our results more efficiently than before. We think our revisions with respect to our point about predictive validity will also help illuminate why our approach provides insight into the accuracy of theoretical explanations. Especially important is the point made on pages 2-3 and 9-10: predictive power is an important criterion for judging any concept to be an important cause of any object of explanation, including state repression. Our approach is valuable for adjudicating between theoretical explanations primarily because it allows us to assess the predictive validity added to models of state repression by measures of concepts theorized to be important causes of state repression.

5) Reviewer 1 writes: The authors state that they cannot think of a reason why democracy would be related to political imprisonment but this seems generally consistent with democratic attempts to hide political repression discussed by Foucault as well as Rejali.

**Response:**

Here we believe Reviewer 1 refers to Foucault's *Discipline and Punish* and Rejali's arguments about stealth torture techniques in democracies (Rejali 2003, 2007). Our argument about democracy and political imprisonment is distinct from the arguments of Foucault and Rejali. We argue that if one considers political competition to be part of the definition of democracy, then targeting political

opposition with violence reduces a government’s level of democracy by definition. Since political imprisonment necessarily targets government opponents (which is not the case with the torture, killing, and disappearance components of CIRI, as discussed on p. 16, footnote 20), it will certainly be related to indicators of political competition, such as Polity’s “parcomp” scale.

Our reading of Foucault is that states, and particularly (but surely not exclusively) Western democracies,<sup>1</sup> have, over time, developed much more subtle and *seemingly* humane, but still coercive, legal sanctions (and surveillance techniques), and that this is not limited to “disciplinary” measures used against political dissidents, but rather includes measures used against people who exhibit any of a range of behaviors deemed to be socially undesirable. We do not believe this suggests that democracies would be systematically more likely to imprison dissidents than autocracies. Additionally, much of this behavior is unlikely to be picked up by indicators as coarse as CIRI and PTS, which measure the most obvious forms of government violence.

Rejali argues that torture in modern democracies happens in private rather than public (partly echoing Foucault’s point), and tends to use techniques that do not leave permanent scars. This suggests that democracies are more likely to engage in stealth torture than scarring torture, and perhaps that democracies are more likely to use stealth torture than autocracies. However, one would need a measure that distinguishes between stealth and scarring torture to test these claims, and the data used in this study cannot make that distinction. This is an interesting claim that should be examined, and one of the authors is actually pursuing this in other work, but it is beyond the scope of this study.

We would like to thank Reviewer 1 for his/her helpful comments and suggestions.

## Reviewer 2 Comments and Responses

1) Reviewer 2 writes: My main concern is with the contribution of the piece to political science in general. The manuscript does not explain why the study is important to the field at large: why is this problem with the repression literature representative of a problem common to a great number of literatures? What do we learn from this study \*that we did not know\* and that can be applied to other types of studies? Most of my points below are

---

<sup>1</sup>For example, it was quite common in the former Soviet Union for dissidents to be declared insane and detained indefinitely in mental hospitals rather than officially tried and imprisoned for dissent.



subpoints of this main idea and are thus mainly concerns with \*framing the contribution\* rather than critiques of the research itself.

**Response:**

As we discuss above, several of the revisions we have made are intended to broaden the appeal of the manuscript by emphasizing the general usefulness of the methods we employ. We think the manuscript benefits enormously as a result of these changes and we thank Reviewer 2 for these suggestions. We especially emphasize that predictive validity, rather than statistical significance alone, should be evaluated in any modeling exercise. This is now stated explicitly in both the introduction (p. 2) and conclusion (pp. 36-37).

2) Reviewer 2 writes: While these analyses are very new to the repression literature, they are not new in political science or even conflict (e.g., Ward, Greenhill, and Bakke 2010). The authors could go much further in explaining why these particular tests are needed. Why is it important to demonstrate the predictive ability of variables? What do we learn that statistical significance does not tell us? This may seem obvious, but this piece could transform the repression literature if the authors can make a convincing argument that our existing studies are doing a disservice to human rights practices by not addressing this issue. I mean, we all know we should make our models explain as much variance as possible, but most scholars believe they are already doing that. So what is different here, in the concept of predictability, that people are currently missing? The paragraphs at the beginning of section 3 do not go far enough to explain why current methods do not imply generalizability, nor how, in a general sense one can imply generalizability.

**Response:**

As noted above, this point is now discussed at several places in the manuscript. On page 3 we first make the case that predictive validity and generalizability are important criteria for evaluating the accuracy of theoretical explanations: if a model produces poor predictions and/or inconsistent results across different sets of data, this calls into question the veracity of the theory that motivated the model. We have also significantly expanded the beginning Section 3 (pp. 9-11), and we think our points are made much more forcefully as a result. We discuss in more detail why statistical significance should not be the only criteria for evaluating theories, and how going beyond null hypothesis tests is beneficial to the repression literature. We also elaborate more on the issue of over-fitting, and cite relevant literature (from both statistics and conflict studies). To illustrate concretely the point that statistical significance and predictive validity are not

the same we have included coefficient plots for three of the models on page 31, and we compare the coefficient estimate for international war (which is large and significant) to plots of cross-validation results for this variable (p. 22) which indicate that it adds no predictive validity to these models. The figures show clearly that including variables with statistically significant coefficient estimates does not necessarily result in a model that fits the data better.

**3)** Reviewer 2 writes: The authors do a very nice job of explaining what these tests are doing, and how they assess predictive ability (though I still don't quite understand how decision trees are determined by the approach). I'd like to see a discussion of how scholars can use these in their own work. These will only transform the literature so that they are commonly used if they are accessible to scholars. Teach us how to use them, or how to otherwise identify the predictive ability of our empirical models. Must we rely on these authors to run these analyses every 10 years with our newly introduced variables of interest? It would be better to demonstrate what scholars should do to demonstrate the importance of their work \*in each study\*. So how should we apply this to our own work?

**Response:**

We have added a paragraph to the conclusion (p. 36) explaining how scholars can incorporate the techniques used in this study into their analyses. Specifically, we urge scholars in all subfields to check model fit if this is not already common practice, which might mean simply evaluating something like RMSE from a linear model, or likelihood based statistics like AIC for models estimated via ML, and seeing if the inclusion of a particular covariate improves these fit statistics. Of course, checking in-sample fit does not guard against over-fitting, so we also encourage researchers to familiarize themselves with cross-validation techniques as these are intuitive, fit nicely within the statistical practice of regression modeling that political scientists are used to, and are relatively easy to implement. On page 36 we cite a working paper by Lee and Ahlquist that is intended to present these techniques to political scientists in an accessible way.

**Reviewer 2 writes:** Please do not assume we all remember what "parcomp" represents as a subcomponent of Polity scores. Always use words when discussing variables.

**Response:**

We agree that using variable labels is unnecessarily confusing, and have changed the manuscript accordingly. We now refer to this variable as either "the competition component of Polity" or "the political competition component of Polity."

We would like to thank Reviewer 2 for his/her helpful comments and suggestions.

## Reviewer 3 Comments and Responses

1): Reviewer 3 writes: The minor revision I suggest is to present the RF method in a clearer and less "jargony" manner. This is a new method for political science, yet - in general - it is relatively straightforward and intuitive. I believe researchers and APSR readers would benefit from a more "gentle" overview and introduction to this method. This should also promote a better overall understanding of your paper and quicker uptake of the method by interested researchers.

I recommend looking at some of the language that will be more familiar to social scientists used in the following paper to describe RF - Stroble, Carolin, James Malley and Gerhard Tutz. 2009. "An Introduction to Recursive Partitioning: Rationale, Application and characteristics of Classification and Regression Trees, Bagging and Random Forests." *Psychological Methods* 14(4): 323-348.

### **Response:**

Our explanation of decision trees and random forests has been significantly improved as a result of the reviewer's suggestion. We have added a cite to the paper mentioned by Reviewer 3, and have attempted to explain the intuition behind these methods in a similar manner. We have expanded the verbal description of both decision trees and random forests (pp. 12-15), and added a brief outline of the procedure on page 12 that we believe will make the discussion easier to follow. We also include a small empirical example using some of the data employed in the paper, and illustrate how decision trees work with a detailed figure on page 14. We have attempted to remove all unnecessary jargon and tried to make the explanation more user-friendly.

We would thank Reviewer 3 for his/her helpful comments and suggestions.

## References

- Rejali, Darius. 2003. “Modern torture as a civic marker: Solving a global anxiety with a new political technology.” *Journal of Human Rights* 2(2):153–171.
- Rejali, Darius. 2007. *Torture and Democracy*. Princeton, New Jersey: Princeton University Press.