THE AMERICAN
POLITICAL SCIENCE ASSOCIATION

APSA – CD

COMPARATIVE DEMOCRATIZATION

"THE PROBLEM OF MEASURING DEMOCRACY"

June 15 - 2017 NEWSLETTER Volume. 17

The American Political Science Association

APSA CD

Volume 15, No. 2 Comparative Democratization June 2017

In This Issue

"The Problem of Measuring Democracy"

EDITORIAL NOTE

Staffan I. Lindberg, V-Dem Institute, University of Gothenburg

Dear Friends and Colleagues,

Now that semesters with teaching and grading are over most places, and summer term is here it may be particularly timely with an issue focusing on how we find out things about democracy, rather than what we have found. We had a similar symposium back in 2011 and it is probably time to revisit the area of measuring democracy. A lot has happened in the last six years. So I am grateful that Steven Wilson agreed to be the issue editor and put together a great collection of contributors for this issue. Thanks to everyone.

While we have an active editorial board, I would also like to remind you that we welcome everyone to approach us with suggestions for symposia for the Newsletter. It is an important forum and your thoughts on what we should bring up for discussion here, are really important. So if you want to take the lead as issue editor on a topic, just let me know.

Finally, this issue is the first with a new Managing Editor. I would like to extend a warm welcome on board the team to Anton Kimfors, student of software engineering at Chalmers University of Technology here in Gothenburg.

ABOUT THIS ISSUE

Steven Lloyd Wilson, V-Dem Institute, University of Gothenburg

In 2017, political science has found a new relevance. To reference the old cliché, we have been cursed with living in interesting times. Comparative democratization in particular has acquired not just a salience, but an immediacy that has not been apparent in a generation, since the heyday of the third wave. Measurement of democracy has become an urgent problem as the resurgence of specters we thought banished to the past call into question just how consolidated our consolidated democracies really are.

This brings to a head the measurement consternation of the long-running debate over whether democracy should be measured dichotomously or along a gradient. In the literature on definitions and measurement of democracy there is substantial variation on the delineation of necessary and sufficient conditions, ranging from Dahl's various formulations of either 5 or 7 requirements for democracy (though he conditions that meeting all of them completely is utopian), to Schmitter and Karl's expansion of Dahl's requirements to 9, to the minimalist point of view of Huntington that focuses almost exclusively on the repetition of free elections.[1] With such disagreement over where exactly the line is between democracy and non-democracy, it is hardly surprising that so many cases end up in a gray area. The so-called "democracies with adjectives" produce an Anna Karenina effect, in which the consolidated democracies tend to be democratic in the same way, whereas autocracies are undemocratic in many different ways.

During the late nineteenth century and the first half of the twentieth, several generations of mathematicians constructed increasingly massive treatises attempting to produce a complete and consistent set of mathematical axioms. The effort was irrevocably demolished in 1931 with the publication of a two page paper by Kurt Gödel that proved that any mathematical system that is consistent can never be complete. During the same time period, political philosophers spent a great deal of time deriving ever more perfect abstract systems of voting in order to guarantee an absolutely fair distribution of legislative seats and determination of executive winners. The key was in the inclusion of ever more layers of preference specification by the voters and ever more complicated schemes of preference aggregation in order to determine the best approximation of the overall social preference. In a publication echoing Gödel's, Ken Arrow's 1951 PhD thesis formally demonstrated that a perfectly fair (in formal terms) democratic system was impossible without requiring at least some element of non-democracy. The result is significant not just for its formal implications but for the more informal conclusion: sometimes the only way to maintain democracy is through non-democratic means. Or as Sartori noted "What makes democracy possible should not be mixed up with what makes democracy more democratic." [2]

Democracy, as Di Palma notes, is "less difficult than we think," arguing that rather than depending on a horde of preconditions, democracy is simply the art of learning to lose.[3] It might

1. Dahl, Robert, Democracy and Its Critics. 1989. Karl, Terri Lynn, and Phillipe C. Schmitter. 1991. What Democracy Is . . . and Is Not. Journal of Democracy 2 (Summer 1991): 75-88. Huntington, Samuel P. 1991. *The Third Wave:* Democratization in the Late Twentieth Century. Norman: University of Oklahoma Press.

2. Sartori, Giovanni. Democratic Theory. (1987)

3. Di Palma, Giuseppe To Craft Democracies. 1990

be a difficult lesson, but it is not a complex lesson. In formal terms, democracy is an infinitely repeating prisoner's dilemma, in which every single turn the losing parties choose to cooperate instead of defecting because the chance of winning in the future trumps the immediate gain of breaking the democratic bargain. But the richness emerges from the informal implications embedded in that logic: democracy depends on whether people trust their political enemies with the machinery of state, because otherwise they would never give up that terrible power. But measuring that trust, that willingness to lose, is problematic. It is a latent concept that we strive to capture by identifying the observable implications of that underlying trust, of elections and previous turnovers of power and the presence of democratic institutions. But in a sense, any measure of democracy in the present is a measure of the probability that the democratic bargain will be respected in the future, in the next turn of the prisoner's dilemma. All democracies are Schrödinger democracies, in that we can never really know if democracy is alive or dead until the electoral box is opened.

This newsletter issue gathers the thoughts of several leading scholars on the problem (and possible solutions) to measuring democracy.

From the perspective of large-n cross national approaches to the problem, Monty Marshall presents a brief history of the development of the Polity series, its methodologies, and its contribution to measuring democracy.

Fabrice Lehoucq addresses the importance of accuracy in our cross-country data sets, arguing that the metrics of consistency among such databases can mask systematic inaccuracies. He takes a deep dive into the state of the art datasets on coups, demonstrating the need for expert understanding of the deep history and context of our cases, even when working in a large-n frame.

Kyle Marquardt presents the Varieties of Democracy Project's approach to aggregating expert coding into large-n measures of democracy, comparing and contrasting with other traditional cross country datasets.

Next, Rasmus Fonnesbaek Andersen argues that the key to measuring democracy in the age of Trump is a return to measurements of political culture.

Ruth Carlitz, Adam Harris, Kristen Kao, Pierre Landry, Ellen Lust, and Lise Rakner present a new approach to conceptualizing governance, focused on a holistic approach that highlights the importance of subnational and community driven understandings of the variables of interest.

Finally, Joshua Tucker presents a thought experiment in article form, theorizing on how political scientists could develop new techniques to use social media data to assess levels of democracy in a number of ways, both direct and indirect.

While this forum is necessarily brief, these articles provide a broad and delightfully varied set of insights into the challenges of measurement endemic to comparative democratization research.

Measuring Qualities of Governance in Complex Societal-Systems: the Polity Series

Monty G. Marshall, Director, Polity Project

Many scholars view "democracy" as an "essentially contested concept."[1] This reference is, at once, inaccurate and appropriate. It is inaccurate because, although highly dynamic and complex, the multi-faceted concept of democracy is knowable and measurable, although its many facets give rise to seemingly endless debates and disputes among those attempting to study or apply its precepts. It is appropriate because the essence of democracy is contestation and open contestation is what most clearly distinguishes democracy from autocracy as a form of governance even as contestation (or conflict) is a "constant" across all forms of governance. Autocracy is known by its use of force, or the threat of force, to repress conflict and compress the basic complexity of politics to a singular, or unilateral, point-of-view. As democracy allows, and is energized by, contestation, it distinguishes itself from autocratic rule through the fundamental rejection of the use of force in addressing, or managing, conflict. Thus, autocracy and democracy can be seen as distinct forms of strategic conflict management: one embracing and the other rejecting the (arbitrary) use of force in maintaining social order and resolving (internal) disputes. This distinction provides the foundation for the Polity™ scheme for measuring contrasting and competing "patterns of authority" so as to better understand "political regime characteristics and transitions."

The Polity data series is one of the most well-known and ubiquitous resources in the social sciences. There are many reasons for its utility and longevity. First and foremost, the Polity data series stands as a macro-comparative record of regime governance characteristics, and changes in those characteristics, based on observed patterns of political behavior involving the complex interactions between state and non-state actors in the formulation and implementation of public policies. As such, it captures essential information on the core object of political science: the state and its governance

1. David Collier, Fernando Daniel Hidalgo, and Andra Olivia Maciuceanu, "Essentially Contested Concepts: Debates and Applications," Journal of Political Ideologies 11.3 (2006): 211-246.

(click to continue on page 8)

Validity is Not Conformity: Measuring Accuracy in Coup Datasets

Fabrice Lehoucq, University of North Carolina

Studying coups is becoming fashionable again. Their prevalence is one reason. Another is the availability of cross-national datasets. But accessibility comes at a cost: we no longer ask many questions about the quality of these coup catalogs. Instead, we argue that consistency among databases demonstrates validity. If something like correlation coefficients among lists are high, researchers conclude, our coup catalog is valid.

But database validity is more than conformity between two or more databases. Its accuracy is another, arguably more central, characteristic of its validity. And accuracy depends upon whether a database avoids falsely claiming something happened ("false positives") or did not happen ("false negatives"). Conformity between two or more databases, in fact, conflates a measure of reliability – to what extent agreement exists among repeated measures of something – with a measure of a database's validity or accuracy.

This paper shows that commonly used databases wrongly claim that a coup – the extra-constitutional overthrow of a chief executive – occurred or did not occur between 19 and 36 percent of the time in 18 countries since 1950. Twenty-eight percent of the coups, for example, in the Center for Systemic Peace (CSP) – the Polity IV – database are mistakes. Most errors consist of failing to note that the military overthrew the chief executive. Mistakes, as I show, are not the product of differences in defining coups. I draw upon a new database that identifies 71 occasions when the military overthrew the president between 1950 and 2016 in 18 Latin American countries (which itself is part of a larger database about 340 coups since 1900, 45 percent of which are successful).[1]

1. Monty Marshall and Donna Ramsey Marshall, "Coup d'état Events, 1946-2015," Center for Systemic Peace (May 11, 2006), http://www.systemicpeace.org/

(click to continue on page 13)

Measuring democracy with expert-coded data

Kyle L. Marquardt, V-Dem Institute, University of Gothenburg

Many important concepts in social-scientific research—perhaps none more so than democracy— are inherently subjective and difficult to measure. Recent attempts to quantitatively measure democracy have used latent variable analyses to square this theoretical circle, arguing that existing measures of democracy are reflections of the same underlying concept.[1] An alternate approach to measuring democracy is bottom-up, i.e. aggregating variables related to the concept of democracy that do not claim to measure democracy itself.

While a dearth of cross-national longitudinal data regarding key concepts related to democracy previously made this latter approach impossible, new datasets—such as those of the Electoral Integrity Project, the Quality of Government Institute and the Varieties of Democracy (V–Dem) Project—have created an opportunity for such research.[2] However, these datasets rely on expert-coded data, which presents new concerns. The key assumption of expert-coded datasets is that the consensus opinion of experts represents a reasonable approximation of a given concept's "true value." Since experts disagree, incorporating dis- agreement into the measurement of a concept is a necessary step in creating expert-coded datasets, and a step which precedes aggregating low-level variables into higher level concepts such as democracy.

In this article, I detail the procedures that one such data gathering enterprise—V–Dem— uses to aggregate its expert-coded data, both at the level of individual variables and high- level indices of concepts related to democracy.

1. Daniel Pemstein, Stephen A Meserve, and James Melton. "Democratic compromise: A latent variable analysis of ten measures of regime type." Political Analysis 18.4 (2010), pp. 426–449; Shawn Treier and Simon Jackman. "Democracy as a latent variable." American Journal of Political Science 52.1 (2008), pp. 201–217

2. Jan Teorell et al. The Quality of Government Standard Dataset, version Jan17. University of Gothen- burg: The Quality of Government Institute, 2017. url: http : / / www . qog . pol . gu . se; Pippa Norris and Max Gro¨ mping. Perceptions of Electoral Integrity, (PEI-5.0). Harvard Dataverse, 2017. url: http: //dx.doi.org/10.7910/DVN/KI5WB4; Michael Coppedge et al. V–Dem [Country-Year/Country-Date] Dataset v6.2. Varieties of Democracy (V–Dem) Project, 2016

(click to continue on page 17)

Assessing and Measuring Political Culture in the Trump Era

Rasmus Fonnesbaek Andersen, University of Copenhagen

Political culturalists claim that societies cannot be ruled wholly without consent, and consequently that citizen preferences and political behavior shape regime type. After Donald Trump's election, a great number of political scientists, political economists and social scientists appear to attribute a key role in preventing democratic breakdown in the US to maintaining democratic norms and values and to civil society resistance. In the case of Daron Acemoglu, his January 2017 piece "We Are The Last Defense Against Trump" in Foreign Policy even suggests a departure from his earlier, vocal skepticism towards political culture.

But political scientists do not usually view political culture as a leading cause of democratic transitions, breakdowns and consolidation or as cause of authoritarian regime stability. The most prominent accounts of democratization, democratic breakdown and regime change today give us little reason to believe that faltering mass support for democracy bodes ill for its survival and spread, or do not give an important place to questions of the perceived legitimacy of and demand for liberal democracy.[1] They posit that political culture is instead endogenous to regime incentives, dynamics and politics, if not irrelevant.

Thus, the focus on democratic norms among the elite and population displayed by political scientists in the media,

1. See amongst others Barbara Geddes, "What Do We Know About Democratization After Twenty Years?" Annual Review of Political Science, 2 (1) 1999: 115–44; Adam Przeworski et al., Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990 (New York: Cambridge University Press, 2000); Carles Boix, Democracy and Redistribution (New York: Cambridge University Press, 2003); Daron Acemoglu & James Robinson, Economic Origins of Dictatorship and Democracy (New York: Cambridge University Press, 2006); Jennifer Gandhi, Political Institutions under Dictatorship (New York: Cambridge University Press, 2008); Milan W. Svolik, The Politics of Authoritarian Rule (Cambridge: Cambridge University Press, 2012); Naunihal Singh, Seizing Power: The Strategic Logic of Military Coups (Baltimore: Johns Hopkins University Press 2014); partial exceptions are Beatriz Magaloni, Voting for Autocracy: Hegemonic Party Survival and Its Demise in Mexico (New York: Cambridge University Press, 2006) and Steven Levitsky & Lucan Way, Competitive Authoritarianism: Hybrid Regimes in the Post-Cold War Era (New York: Cambridge University Press, 2010).

(click to continue on page 22)

Beyond the State: Measuring Governance at the Community Level

Ruth Carlitz, University of Gothenburg
Kristen Kao, University of Gothenburg
Ellen Lust, University of Gothenburg
Adam Harris, University of Gothenburg
Pierre Landry, New York University
Lise Rakner, University of Bergen

The international development community's interest in "good governance" has been growing steadily since the 1990s, when development practitioners came to realize that initiatives which looked good on paper frequently failed due to gaps in policy implementation. This realization led international financial institutions such as the World Bank and the International Monetary Fund to focus on governance reforms as a linchpin of their development assistance programs. At the same time, social scientists were becoming increasingly dissatisfied with their inability to describe and explain the "real world." Such dissatisfaction motivated a number of disciplinary reorientations, including the rise of new institutional economics. In shifting their focus to institutions, economists paved the way for political scientists to develop new research agendas examining the form and functioning of state institutions, and the implications of institutional variation for a range of outcomes.

Despite this surge of interest, scholars and policymakers have yet to reach a consensus regarding the definition and measurement of governance. Current measures tend to conceive of governance as a national-level phenomenon and focus exclusively on formal political institutions. However, the nature and quality of governance can vary across both subnational geographic areas as well as within different "imagined" communities that extend beyond geographic boundaries. These include communities based on shared social identities such as ethnicity, class, and gender. The extant

(click to continue on page 26)

Can We Measure Democracy Using Social Media?

Joshua A Tucker, New York University

The purpose of this forum is to consider alternative "outside the box" ways of thinking about how to measure democracy. Social media seems like an obvious candidate for this task. After all, people are currently trying to use social media to measure all sorts of things,[1] and there are in fact enormous quantities of social media data currently in the public domain that record the musings of billions of people around the world. Moreover, social science has been busy over the past decade developing all sorts of tools to turn text into data, thus making the possibility of harnessing social media for quantitative analysis that much more tantalizing, even beyond the fact that social media "posts" are often accompanied by all sort of metadata that already is perfectly suitable for quantitative analysis.

That being said, I am not currently aware of any study that purports to use social media data to provide a "direct" measure of quality of democracy. In the remainder of this essay, I will therefore explain why exactly this would be so challenging – albeit with tantalizing opportunities as well – but in addition propose three alternative ways that social media data could be used to try to indirectly get at the strength of democracy in a country: by measuring the presence (or lack thereof) of mass attitudes associated with support for democracy;[2] by using social media as a tool to measure other indicators that are thought to be correlated with democracy (e.g., tolerance); and finally by assessing patterns of

1. To name just a few related to comparative politics, scholars are using social media data to try to measure political polarization (Barberá et al 2015; Barberá 2016), information diffusion in protests (Gonzalez-Baillon et al 2011; Barberá et al 2015; the prevalence and spread of fake news (Allcott and Gentzkow 2017); communication of foreign policy positions (Zeitzoff, Kelly, and Lotan 2015); censorship priorities of authoritarian regimes (King, Pan, and Roberts 2013, 2014); communications strategies of both opposition and pro-regime elites during periods of protest (Munger et al.), etc.
2. See for example Linz and Stepan (1996), who define democracy in part as being "the only game in town" in the minds of the mass public.

(click to continue on page 30)

Marshall, CONTINUED

(continued from page 5)

regime. The principal strength of Polity data derives from its macro-comparative basis: the project uses standardized rules for coding general (institutionalized) patterns of political behaviors that are systematically applied to all units in the universe of analysis (in this case, all independent countries with a total population greater than 500,000; 167 countries in 2016) over a substantial period of time (1800 to present). The availability of standardized, comparable measures of key quality of governance indicators across units of analysis over time allows for the macro-comparative analysis of the interaction between varying qualities of regime governance and policy-related inputs or outcomes of interest. Macro-comparative analysis provides an empirical basis for assessing the nature of conditional and/or processual relationships among attributes in the complex, interactive schemes that characterize modern societal-systems. The policies of central governance condition all political actions in a political system, either directly through setting the political agenda or indirectly by defining the nature of opposition to the system. The general characteristics, or "qualities," of the state's practical authority must be taken into account in any analysis of the effects of political behavior. For example, one of the most robust findings drawn from analyses using Polity data is that "hybrid" or "anocratic" regimes (those with some mixture of autocratic and democratic authority traits) have higher risk of political instability.

Another core strength of the Polity series lies in its conceptual evolution and refinement over time. Like most macro-comparative data collections, Polity came into existence with limited resources as a result of the personal research interests of its creator; initial, "rough-cut" data collections are conducted to investigate a particular research question and, in doing so, gauge the feasibility, veracity, and utility of the data to support a potentially larger and more refined research agenda. Specifically, the Polity scheme grew out of the prescient interest of Ted Robert Gurr in the systemic basis for rebellion by anti-state groups which he first articulated in his 1970 book, Why Men Rebel. In addition to bringing state and societal forces together into the analysis of political outcomes, Gurr was also a leading proponent of a new approach to empirical investigation to coincide with the advent of computer processing and the beginning of the "behavioral revolution" in the social sciences: macro-comparative analysis, which he termed "politimetrics" (relying on "coded" indicators) and stood in contrast with econometric accounting and statistical methods (using "counted" indicators). Political behaviors are complex amalgams of large numbers of inter-dependent, individual actions that are more-or-less disciplined through the direction of group leadership and institutionalized authority. The design, or "operationalization," of clearly articulated rule sets for coding key, observable characteristics defined by collective behaviors for use in macro-comparative analyses may be as close as researchers can get to scientific objectivity in the social sciences. [2]

Social science researchers in the early 1970s were also grappling with the dismantling of colonialism and the emergence of "nativized" central authority structures in newly independent countries across Asia and throughout Africa. The diversity of societies in this greatly expanded universe of states argues against the oversimplification of complex, social phenomena into "ideal types," even though those idealized typologies may be considered "preferred states." Regardless of intent, ideal types prove useful by providing standards for measuring "distance" from the ideal "pole." Dichotomized "ideal" representations of complex, social phenomena such as "autocracy" and "democracy" hold very little information beyond a base distinction that can only be analytically useful for making gross assessments with little or no explanatory potential. The newly independent states of the post-colonial period, 1955-1975, lacked social cohesion and institutionalized political practices and their inherent volatility often resulted in erratic movements between autocratic and democratic authority practices.

2. Ted Robert Gurr, Why Men Rebel (Princeton University Press, 1970); Ted Robert Gurr, Politimetrics: An Introduction to Quantitative Macropolitics (Englewood Cliffs, NJ: Prentice-Hall, 1972; Ted Robert Gurr, "Persistence and Change in Political Systems, 1800-1971," American Political Science Review 68 (1974): 1482-1504; and Harry Eckstein and Ted Robert Gurr, Patterns of Authority: A Structural Basis for Political Inquiry (New York: John Wiley & Sons, 1975).

The Polity scheme was designed to map the space between the autocratic and democratic poles to better capture the volatility of transitional politics and instability of "non-institutionalized" regime structures. The Polity scheme expands on potential analytic capacity by disaggregating information contained in the two generally-recognized, ideal regime types: autocracy and democracy, into 218 operant permutations, that is, distinct, extant combinations of values recorded in the six component variables (or three "concept variables," see below) in the Polity scheme: executive recruitment (8 combinations), executive constraints (7 combinations), and political competition (10 combinations). Despite its broad analytic potential, most researchers appear to have preferred using the 21-point POLITY or POLITY2 "scales" which combine the separate measures of autocratic (AUTOC) and democratic authority (DEMOC) in a single scalar measure. Of course, the question of "interval" when using the POLITY or POLITY2 scalar variables looms large in analyses as the assumption of standardized "distance" between points along the scale(s) is(are), at best, an approximation that conflates distinct permutations in constructing a scalar array and, so, reduces (218) operant permutations by nearly a factor of ten (21).

As noted above, "living" data resources, that is, ones that are maintained, refined, and updated regularly over time, must be useful, comprehensible, and trusted by researchers, and research consumers. Thus established, Polity has become ubiquitous in quantitative research in political science, therefore, it can serve as a common point or nexus among many research streams and as a lingua franca in the process of knowledge synthesis and accumulation, just as the state links a panoply of political interests and agendas in formulating common policies. Living data resources, in order to foster support over time, must adapt and evolve in response to changing technologies, circumstances, and the research needs of a broad and active consumer base. Polity can be considered a maturing data series. Polity has been conceptually refined during each of its four development phases. The first three phases took place under the direction of Ted Robert Gurr. The original Polity data collection was one of four suites of data compiled to investigate the nature of civil strife and the magnitude of civil conflict over the period 1800-1971; the four suites detailed Civil Strife Events, Conflict Magnitudes, Conflict and Society, and Polity Persistence and Change. It was designed to examine the relationship between civil conflict and regime authority to better understand the nature of political stability and regime durability. The original study comprised an enormous, ten-year effort; the data structure was strictly comparative: the "polity" was the unit of analysis and the examination focused on the origination and termination of that "polity" (the Polity suite, itself, contained 85 points of information on each "polity" case). The second phase in the data development, Polity II, updated the coverage through 1986 and shifted the unit focus from the "polity" or "regime" to the "country"; the data format was restructured to an annual time-series to make it consistent with the prevalent standard for quantitative analysis in political science and compatible with other data resources for multivariate analyses. The data core of the Polity II iteration comprised nine "basic authority characteristics" and two derivative indices: democracy and autocracy scores; regime changes in a particular country had to be accounted for, so, indicators of the nature of regime changes were also included. The third phase, Polity III, updated the data through 1994 and streamlined the collection to include only the basic authority characteristics and the two derivative indices of democracy and autocracy. The first three phases also included measures of political "power" and the concentration of that "power" in the central state. The Polity Project remained a largely academic enterprise and derived its principal support from the home institutions of its principal investigator and grants from the National Science Foundation. In the mid-1990s, a supplemental version of the Polity III data series was developed by researchers at the Peace Research Institute of Oslo (PRIO) under the direction of Nils Petter Gleditsch that searched through the original Polity coding sheets in order to date all changes in Polity scores and reformat the data back to the original polity-case

listing (version Polity IIId).[3]

By 1998, the Polity measures of regime governance had become a key input in a global modeling effort for forecasting political instability supported by the US Government (USG) intelligence community and conducted under the auspices of the State Failure Task Force (created in 1994 and now know as the Political Instability Task Force, PITF). To this writing, Polity continues to draw primary support from the Task Force. Direction of the Polity Project was transferred to the author in 1998. The first update of the Polity data under the aegis of the PITF (version Polity98) brought the coverage through 1998 and stripped the series down to its core: the six component variables (regulation, XRREG, competitiveness, XRCOMP, and openness, XROPEN, of executive recruitment; executive constraints, XCONST; and regulation, PARREG, and competitiveness, PARCOMP, of political participation) and the two indices derived from those variables: democracy (DEMOC) and autocracy (AUTOC) scores ("power" variables were dropped). Polity's integration with the PITF modeling effort had several profound effects on the data series. Polity's strict reliance on systematic observation of political behaviors , that is, practical politics, made it compatible with the policy-oriented perspectives of practitioners, both intelligence analysts an policymakers. However, it had difficulty speaking their "language." Conceptual abstractions, taken from historical, observed behaviors and tossed into an analytic "soup" with other conceptual abstractions must yield "findings" that can be interpreted and translated back into relatable, actionable applications to complex, unfolding situations of immediate interest, a process this author terms "re-contextualization." This is not generally considered a requirement in academic research. The data inputs and model findings also had to be explainable in credible, if not compelling, narratives that are useful and able to withstand increased, direct, and detailed scrutiny by knowledgeable experts. In order to be accessible to intelligence consumers and a policy audience, Polity had to be substantially enhanced; it also needed to be updated more regularly.

A simple factor analysis of the Polity data showed that there were eight combinations of the values coded for the three nominal variables contributing to the definitive form of executive recruitment (EXREC); seven ordinal levels of institutional "checks and balances" or constraints on executive action (EXCONST), and ten combinations of the values on the two nominal variables defining the basic forms of political competition (POLCOMP) involving the broader public constituency. Unlike the largely nominal, component variables informing the original data series, the extant combinations of related values appeared to hold an intrinsic ordinality along a spectrum stretching between autocratic and democratic authority poles, constituting increments of change in a transition back-and-forth from one pole to the other. The resulting three "conceptual variables" are more intuitive and, so, easier to apply and interpret. In addition, increased levels of scrutiny demand greater transparency so that users and consumers have a better understanding of the coded information and how it relates to the broader context from which it has been distilled. Because the Polity series has an annual time-series structure, more precise information on the timing of changes in the data records were needed; having the time record incorporated in the data series means that the data format can be converted a monthly, or even daily, time-series. The fourth phase of the Polity series, Polity IV, then, is monitored daily, reported monthly, and updated annually; it includes the three, combined "concept" variables, a measure of regime durability (time since the most recent, substantive change in Polity codes); specific dating of coded changes; and measures of the degree of all Polity changes. Individual country reports were written to provide a narrative description of the circumstances coded by Polity; however, these quickly proved too costly to update on a annual timetable. The most important enhancement to the Polity IV data series is greater accuracy resulting from the availability of better information on political behaviors, especially since the end of the Cold

3. Copies of the first three phases of the Polity data series are deposited with the Inter-University Consortium for Political and Social Research (ICPSR; www.icpsr.umich.edu). For a report on the Polity IIId project, see, Sara McLaughlin, Scott Gates, Håvard Hegre, Ranveig Gissinger, and Nils Petter Gleditsch, "Timin the Changes in Political Structures," Journal of Conflict Resolution 42 (1998): 231-242.

War and the attendant "national security" restrictions on reporting the internal affairs of states.[4]

As the PITF continuously refined and enhanced its modeling of the pre-conditions of political instability, pressures to refine and enhance model inputs continually intensified. Changes have had to be monitored in near real-time, meaning that an immediate comprehension of relevant changes in any of, now, 167 countries of the world has had to be maintained. As questions have arisen regarding particular codes for particular cases, those cases have been re-examined and revised when appropriate; greater utilization and scrutiny of the data means that even random data errors are discovered and corrected. Most importantly, the centrality of the Polity variables in the modeling effort and the greater analytic pressure placed on the research to advance forecasting goals led to a systematic "drilling down" into the foundations of the data resources in order to discover the "explanatory core" of the Polity data inputs. Most prior research using Polity, whether by the PITF or the broader research community, had focused on the derivatives, mainly the combined POLITY index (DEMOC–AUTOC) used either as a scaled measure or as a basis for categorizing regime types. Testing the foundations of the Polity data, that is, the six component variables, raises the analytic potential of the data series while pushing for even greater refinement of, and consistency in, the data. PITF refinement of its global model revealed that the most powerful explanatory indicator of impending political instability is the condition coded in the Polity data as "factionalism" (i.e., PARCOMP=3). [5] This finding triggered a five-year investigation, led by this author, into the validity of that coding and the veracity of the coded behavior. The investigation confirmed the validity and veracity of the "factional" coding and led to a more detailed understanding of the condition and further insights into the dynamics of that condition. As James Vreeland pointed out in his 2008 critique, "factionalism" is a conflict condition that can lead to an onset of civil war; however, what the PITF research found was that "factionalism" was far more likely to lead to an "adverse regime change," or "autocratic backsliding" (that could then precipitate an onset of civil war). [6]Vreeland's "conflict contamination" criticism zeros in on the core of the Polity scheme. Polity measures the conflict management strategies and capabilities (authority) of the particular governance regime. Two essential conflict management strategies are available to political leaders: use of "instrumental" strategies, or coercion, preferred by exclusive regimes (autocratic authority: rule by force) and "sociational" (cooperative) strategies preferred by inclusive and integrated regimes (democratic authority: rule of law). Electoral procedures politicize the system and reveal its authority characteristics, but it is the general consensus underpinning sociational strategies not to use force in the resolution of conflicts that determines democratic authority. Conflict management is the core function of governance in complex, societal-systems.

The fifth phase of the Polity series, Polity V builds upon the detailed investigations conducted in regard to the "factional" finding. The "factional" coding is found in the Polity records for nearly two-thirds of the countries covered by the data during the contemporary period (since 1946). The first thing discovered during this investigation was that the dating of Polity changes, done by the PRIO researchers using

4. This research was conducted for the Political Instability Task Force which is funded by the Central Intelligence Agency. The views expressed herein are the author's alone and do not necessarily represent the views of the Task Force or the U.S. Government. Because of the USG support, the Polity data series and its supporting documents are required to be freely available to the public; they are posted without delay to the INSCR Data page on the Center for Systemic Peace web site (www.systemicpeace.org/inscrdata.html). A log of all changes made to the series during the annual update cycle is also posted there.

5. "Factional" is defined by Polity as "Polities with parochial or ethnic-based political factions [including those that control the state] that regularly compete for political influence in order to promote particularist agendas and favor group members to the detriment of common, secular, or cross-cutting agendas." Marshall and Cole 2008 (note 6) more accurately term this condition "polar factionalism" as it almost always involves a polarization of societal groups into two opposing factions.

6. Monty G. Marshall and Benjamin R. Cole, "A Macrocomparative Analysis of the Problem of Factionalism in Emerging Democracies," paper delivered at the 2008 Annual Meeting of the American Political Science Association; Jack Goldstone, Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael Lustik, Monty G. Marshall, Jay Ulfelder, and Mark Woodward, "A Global Model for Forecasting Political Instability." American Journal of Political Science 50.1 (2010): 190-208. James Raymond Vreeland, "The Effect of Political Regime on Civil War: Unpacking Anocracy," Journal of Conflict Resolution 52.3 (2008): 401-425.

the original Polity coding sheets, was not as precise as hoped; many of the dates were simply estimates. We also found some inconsistencies in the application of coding rules and conventions for certain conditions affecting particular component variables, notably "executive-guided transitions" (EXREC=5); weak political party structures (POLCOMP=4 or 5); and the those concerned with the quality of election processes (POLCOMP=4 or 5; EXREC=7 or 8; and POLCOMP=8, 9, or10). While these inconsistencies do not substantially affect the aggregate indices (AUTOC, DEMOC, and POLITY), they do limit the potential for more detailed analyses based on the Polity component variables; they also present a bias moving forward, as the world is becoming more democratic and these "quality of participation" effects mainly impact the democratic end of the POLITY "spectrum." Polity V, then, is using higher quality and more detailed information resources to conduct systematic reviews of political behaviors found in news reports in all countries over the period 1946 to present in order to fine tune all Polity codings, and datings of coding changes, to improve accuracy and consistency. It will also improve transparency by producing chronologies, or process tracings, identifying and explaining all Polity changes since 1946. In doing this, Polity V will gain greater separation at the democracy end of the POLITY spectrum by distinguishing more nuanced qualities of democratic authority patterns. The Polity Project has also added, or will soon add, enhanced information on coups and other types of irregular changes in executive leadership, electoral boycotts, elections, and executive and party structures through the development of "plug-in" data series that can be added to the base Polity data series.

The real strength of the Polity series, and the reason it has endured to serve as a lingua franca for empirical research in the social sciences, is that it taps into a core truth in the structures human beings have developed to manage conflict in their increasingly complex, societal-systems. In the author's own theoretical work, the Polity scheme is recognized as containing the "master key" in understanding the science of conflict management and political behavior, more generally.[7] For most data consumers, data collection activities simply provide a product; one of many resources on which they can draw to help them better understand and illuminate (some part of) the human condition. To the data collector, the data collection process is essentially a disciplinary tool in the systematic process of observation and comprehension of the structures and dynamics of complex, societal-systems, that is, the accumulation of knowledge and the identification and explanation of key patterns. Mapping the patterns that constitute complex systems informs us of how systems work, how they might prosper, and how they can fail.

7. Monty G. Marshall, Third World War: System, Process, and Conflict Dynamics (Lanham, MD: Rowman & Littlefield, 1999); Monty G. Marshall, Managing Complexity in Modern Societal-Systems: Structuration (vol. 1) and Problemation (vol. 2), video book, Eliot Elzinga, videographer (Vienna, VA: Center for Systemic Peace; access www.systemicpeace.org/videobook.html)

Lehoucq, CONTINUED

(continued from page 5)

Validity as Conformity

When comparativists explore the validity of their databases, their reasoning is that compatibility equals validity or, at least, underscores its credibility. Curtiss Bell, for example, mentions that the CSP and Powell and Thyne's datasets are correlated at the 0.844 level. Table 1, which presents partial correlation coefficients for 18 Latin American countries between 1950 and 2015, uncovers unimpressive levels of agreement among lists of coups (again, I ignore coup attempts). Most partial correlation coefficients are less than 0.500. The lists compiled by CSP and Powell and Thyne are correlated at the 0.804 level (again, these and the next coefficients refer to the 18 Latin American cases). But each of these correlate at the 0.364 level with the lists of coups that Marinov and Goemans extract from the Archigos database, which is an important database of executive succession gathered from news reports.[2] And the Archigos database is correlated with the CST at the 0.430 level.

Most lists are not highly correlated with mine. The highest agreement is between my list and Belkin and Shofer's, which is 0.664, and mine and Powell and Thyne's, which is 0.633. But my list is barely compatible with Archigos's (0.203). Agreement between my list and several just on

inscrdata.html, accessed August 1, 2016 and Fabrice Lehoucq, "Database on Military Coups: Dates, Characteristics, and Dynamics," unpub., 2017.

2. Curtis Bell, "Coup d'états and Democracy," Comparative Political Studies, 49:9 (2016): 1167-1200, Goemans, Henk E., Kristian Skrede Gleditsch, and Giacomo Chiozza, 2009, "Introducing Archigos: A Dataset of Political Leaders," Journal of Peace Research, 46:2 (2009): 269-283, Marinov, Nikolay and Hein Goemans, "Coups and Democracy," British Journal of Political Science, 44:4 (2014): 799-825, Monty Marshall and Donna Ramsey Marshall, "Coup d'état Events, 1946-2015," and Jonathan M. Powell and Clayton Thyne, "Global Instances of Coups from 1950 to 2010: A New Dataset," Journal of Peace Research 48:2 (2011): 249-259.

TABLE 1
LOW LEVELS OF AGREEMENT AMONG COUP CATALOGS
(PARTIAL CORRELATIONS, WITH LEVELS OF SIGNIFICANCE AND NUMBER OF CASES)

	Lehoucq 2016 (-2015)	CSP (-2015)	Powell& Thyne (-2015)	Marinov& Goemans (-2004)	Belkin& Shofer (-2000)	Lehoucq 2014 (-2006)	Smith (-2000)	Nohlen (-2000)
Lehoucq	1.0000							
	111							
CSP	**0.5315**	1.0000						
	0.0000							
	111	111						
Powell & Thyne	**0.6448**	**0.8031**	1.0000					
	0.0000	0.0000						
	111	111	111					
Marinov & Goemans	**0.2656**	**0.4458**	**0.3568**	1.0000				
	0.0053	0.0000	0.0001					
	109	109	109	109				
Belkin & Shofer	**0.6762**	**0.8070**	**0.7784**	**0.3531**	1.0000			
	0.0000	0.0000	0.0000	0.0002				
	108	108	108	107	108			
Lehoucq 2014	**0.5650**	**0.4816**	**0.5029**	**0.0490**	**0.5931**	1.0000		
	0.0000	0.0000	0.0000	0.6127	0.0000			
	111	111	111	109	108	111		
Smith	**0.6627**	**0.6311**	**0.6746**	**0.3527**	**0.7230**	**0.6677**	1.0000	
	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000		
	109	109	109	107	108	109	109	
Nohlen	**0.5702**	**0.5099**	**0.5053**	**0.2080**	**0.4639**	**0.6371**	**0.6270**	1.0000
	0.0000	0.0000	0.0000	0.0316	0.0000	0.0000	0.0000	
	108	108	108	107	108	108	108	108

Source: See footnote one.

Latin America do somewhat better. In his Democracy in Latin America, Peter H. Smith combines the results of several compilations, including those by Warren Dean, Martin Needler, and Mauricio Solaún and Michael Quinn. His list correlates with mine at the 0.663 level for the second half of the twentieth century, but at the 0.466 level for the years since 1900 (which I do not display).[3]

It is not differences in definitions that reduces levels of agreement among coup catalogs. The CSP data, according to Marshall and Marshall's codebook, contains information on "…successful, attempted, plotted, and alleged coup events reported in Keesing's Record of World Events (Keesing's Online) and other sources." Jonathan M. Powell and Clayton L. Thyne's dataset, which merges the results of 14 lists of coups (and coup attempts) before searching three electronic news archives, defines a coup (or a coup attempt) (on p. 250) as an effort "undertaken by any elite who is part of the state (emphasis in the original)" to "overthrow the chief executive." Marinov and Goemans's definition is looser, which does help to explain why correlation coefficients are lower. They define a coup (on p. 809) as a situation "where (1) force is threatened or used, (2) the military, government insiders or rebel actors are involved and (foreign) actors play a minor role."

Correlation coefficients, at any rate, are a debatable measure of compatibility for binary data. They express the strength and nature of relationships between linear or at least ordinal data containing more than 2 values. Knowing that list B is highly compatible with list A, moreover, cannot answer the question we should pose: which list should we trust?

Validity as Accuracy

Powell and Thyne's list moves beyond discussing correlations. Instead, it reviews evidence – news archives of key newspapers of the Western (mostly US) press – to establish a benchmark with which to assess 14 coup catalogs. While Powell and Thyne do not emphasize the importance of accuracy, their analysis discloses compatibility with the 227 coups they document between 1950 and 2015. Powell and Thyne exclude what they call false positives from their catalog of coups. These are cases for which they could find evidence of a coup, including because they could not find evidence of anti-government activity or of overt attempts to seize power. There are, in other words, numerous errors of interpretation in existing coup catalogs. Powell and Thyne's efforts, I should note, is part of a larger list that also includes 230 failed attempts; 457 is the total number they cull from a candidate list of 1,259 reports of sedition (from the 14 original lists).

Powell and Thyne find that the average rate of false positives is not low (which they define as the number of these cases divided by this catalog's total number of alleged coups). It is 33.1 percent. The range runs from 16.4 (Morris Janowitz's 1977 list of coups) to 49.1 percent (Moreno's 2004 list of coups in Latin America). They also present a measure of accuracy. It is the product of dividing the number of true positives and true negatives by the total of all allegations of coups (or all true positives and negatives and all false positives and false negatives). The average across the 14 lists is 75.7 percent. The range is from 67.7 (Moreno's list) to 88.3 percent (for William R. Thompson's 1973 coup catalog for all countries between 1946 and 1970; complete citations to these studies are available in Powell and Thyne's article).

Table 2 uses my database as a benchmark to assess the accuracy for 7 coup catalogs for 18 Latin American cases between 1950 and 2015. First, this table discloses that the false positive rate is rather low for most lists. Smith's has the least number of false positives: 2 percent. Marinov and Goeman's list has the highest rate at 28.2 percent. The average rate of false positives is 9.4 percent. And the list that I assembled by 2012 and published in 2014 has a 13.6 false positive rate.[4] Second, accuracy ranges between 64 and 82 percent with my list (the mean is 75.8 percent). The most accurate list is Smith's (and, followed, as it turns out, by Powell and Thyne's at 81 percent). The least accurate is Marinov and Goemans' at 64 percent. My earlier list is 80 percent accurate.

Errors stem largely from failing to record the overthrow of a chief executive. With one or two exceptions (see prior paragraph), most of these lists infrequently make false positive errors. The average rate of correctly reporting a coup is, on average, 72 percent. What accounts for the discrepancies between coup catalogs? Again, with the exceptions of Marinov

3. Aaron Belkin and Evan Schofer, "Toward a Structural Understanding of Coup Risk," Journal of Conflict Resolution, 47: 5 (2003): 594-620, Peter H. Smith, Democracy in Latin America: Political Change in Comparative Perspective (Oxford: Oxford University Press, 2005), p. 354; also, see Warren Dean, "Latin American Golpes and Economic Fluctuations, 1823-1966," Social Science Quarterly, 51: 1 (1970): 70-80, Martin Needler, "Political Development and Military Intervention in Latin America," American Political Science Review, 60: 3 (1966): 616-626, Mauricio Solaún and Michael Quinn, The Politics of Military Intervention in Latin America (Urbana: University of Illinois Press, 1973).

4. Fabrice Lehoucq and Aníbal Pérez-Liñán, "Breaking Out of the Coup Trap: Political Competition and Military Coups in Latin America," Comparative Political Studies, 47:8 (2014): 1105–1129.

TABLE 2
UNIMPRESSIVE LEVELS OF ACCURACY AMONG COUP CATALOGS

	CSP (-2016)	Powell & Thyne (-2016)	Marinov Goemans (-2004)	Belkin & Shofer (-2000)	Lehoucq 2014 (-2006)	Smith (-2000)	Nohlen (-2000)
False Positive Rate*	4.5%	5.4%	28.2%	3.8%	13.6%	2%	8.2%
Accuracy**	71.8%	80.9%	63.6%	80.0%	80.0%	81.8%	72.7%
N	42	53	51	51	57	49	45

*This rate = number of false positive errors/total number of alleged coups by this source.
**This rate = number of true positives and true negatives/by total number of true and false positives and negatives. The benchmark list is Lehoucq 2017 (presented in this paper), which lists 71 coups for this period.

and Goemans' and my earlier effort, most errors consist of false negative reports. But Marinov and Goemans as well as my earlier effort suffer from false positive reports. These sources, in other words, mistakenly label attempts, opposition rebellions, or autogolpes (incumbent takeovers) coups.

A Comprehensive List of Coups

Answering the question of why coup catalogs contain a large error rate has a short answer: researchers have not done their homework. It is a charge to which I also plead guilty; in my 2014 paper, I also spent an insufficient amount of time identifying relevant sources, reading them, and struggling with what can be an incomplete and even inconsistent historical record.

To reconstruct the actual history of coups, my research assistants and I read the more specialized monographs about military governments, dictatorships, political conflicts, regimes, and political succession for each country. We compared the often fragmentary record from these sources to construct a portrait of coups and coup attempts. We also searched the British and US diplomatic files for information about them. For quite a few coups, a handful of sources (often just in Spanish) jointly yield a basic chronology of events, the identity of key military officers and politicians, and enough evidence to understand why sedition fails or succeeds.

Inconsistencies as Accuracy Tests

There are 16 major inconsistencies between my database and the other coup catalogs. Slightly less than 75 percent of these are false negatives. There are 11 cases, in other words, where other catalogs fail to record cases of the military overthrowing the president. An example is the coup in Argentina in 1981. In a detailed monograph on the Argentine military, Robert Potash points out that General Lanusse gained the support of the military to overthrow President (and General) Levingston, whom was resisting the liberalization of the regime that the military junta wanted. Another is the assassination of President Rafael Trujillo of the Dominican Republic in 1961. While I do not label most assassinations coups, Trujillo's assassination was part of a much broader conspiracy by military factions to unseat the long-term dictator. Piero Gleijeses's extensively researched monograph on how instability led to the US invasion of the Dominican Republic in 1965 makes this clear. One of these coups is an error of my own, of claiming a coup occurred in Panama in 1983, when it, in fact, happened a year earlier. [5]

The remaining five cases are false positive errors. Most of these stem from mislabeling unusual changes in executive succession. A poet, Rigoberto López Pérez, assassinated President Anastasio Somoza of Nicaragua in 1956; no evidence exists that members of the National Guard conspired against the dictator. President Jânio Quadros of Brazil resigned in 1961 because he lost a struggle with Congress. The authoritative study of this period by Thomas Skidmore mentions no military conspiracy to topple quixotic Quadros. We also cannot claim a coup happened in Uruguay in 1976 if we record a coup in 1973. One of the most well-known coups in Latin America felled a democracy in 1973 when the

5. Piero Gleijeses, The Dominican Crisis: The 1965 Constitutionalist Revolt and the American Intervention (Baltimore, MD: Johns Hopkins University Press, 1978) and Robert Potash, The Army and Politics in Argentina, 1962-73 (Stanford, CA: Stanford University Press, 1996).

military got President Bordaberry to recognize the supreme authority of the high command, even as officers closed Congress and jailed thousands of citizens. But because Bordaberry remained a figurehead until his term ended in 1976, some lists prefer to label this as the coup year. There was no coup in Panama in 1983. And I mislabeled a coup in Panama in 1981, when it, in fact, happened a year later. [6]

Note that the data in Table 2 corrects these 3 errors in my final list of coups. I mention these efforts to disclose that I compared the results of my research with existing catalogs of coups before (yet again) returning to the relevant historiography to arbitrate between competing claims (and even though most lists provide no documentation to reconstruct their assessment of the evidence). But I list the full extent of the errors of my earlier list. Most of these errors stem from uncritically accepting the results of other coup catalogs. To guard against mistakes, my research assistants and I have written more than 700 pages of text documenting and explaining our decisions.

Implications

The evidence in this brief article should give us pause. In the rush to assess cause and effect about political events like coups, we have sidelined the discussion of accuracy. We have substituted measures of reliability—how compatible two or more portraits of the same phenomenon are—for those of database validity. But validity is more than just conformity.

Two implications follow from this argument. First, validity—accuracy, as I conceive it—requires thorough description, which political scientists have marginalized over the past decades.[7] It requires analyzing all the relevant evidence and evaluating inconsistent and contradictory reports to create verifiable accounts of coups (and attempts). And it requires delineating research protocols to match evidence to clearly defined concepts. It is the large number of steps we take from observation (which involves judging the usefulness and credibility of qualitative reports) to inclusion in datasets, something that relies upon sifting through multiple (and occasionally inconsistent) accounts of military insubordination to judge whether an unconstitutional seizure of power is a coup. Unfortunately, most coup catalogs do little more than mention they use Keesing's or the New York Times. Few identify all of their sources and how they arbitrate among inconsistent or contradictory claims.

An important part of such efforts should include a discussion of problematic cases, which is why I investigate inconsistencies between other databases and my own. Investigating allegations of coups plumbs the depths of our research. Comparing databases gauges the usefulness of our criteria of what is coup. And reconciling claims between databases tests the comprehensiveness of our research protocols.

Second, the soundness of existing research on military coups is in doubt. True, my validity test only encompasses 18 Latin American countries. But, if something like 25 percent of a coup catalog's entries are wrong, then statistical results may disintegrate. And, if a central effort of statistical analysis is to illuminate dynamics of particular cases, then getting the cases right will improve our efforts to improve explanations of the causes and consequences of coups.

6. Knut Walter, The Regime of Anastasio Somoza: 1936-1956 (Chapel Hill, NC: The University of North Carolina Press, 1993), Thomas Skidmore, Politics in Brazil 1930-1964: An Experiment in Democracy, updated ed., (New York: Oxford University Press, 2007), Charles Gillespie, Negotiating Democracy: Politicians and Generals in Uruguay (Cambridge, UK: Cambridge University Press, 1991), and John Dinges, Our Man in Panama: How General Noriega Used the United States – and Made Millions in Drugs and Arms (New York: Random House, 1990)

7. John Gerring,"Mere Description," British Journal of Political Science, 42:4 (2012): 721-46.

Marquardt, CONTINUED

(continued from page 6)

The V-Dem Project relies on over 2,800 experts who each score a subset of the 150 subjective variables in the V–Dem dataset; the dataset includes an additional 128 "factual" variables that a single researcher codes. The project aggregates its low level factual and expert-coded variables into both meso- and high- level indices. The aggregation procedures for these higher level variables rely on a mixture of standard aggregation algorithms and theory-driven approaches. I focus on one of V– Dem's theoretically-derived high-level indices, "Polyarchy," which measures the concept of electoral democracy. I describe both the measure's theoretical basis and operationalization, comparing it to an inductive measure of a similar concept. The comparisons speak to both the advantages and disadvantages of theory-driven aggregation criteria.

1 Aggregating expert-coded data

Expert-coded datasets generally involve codings on an ordinal scale. To summarize the re- sulting data, producers of these datasets often report the mean of the expert-coded values. Such an approach is problematic for several reasons. First, it assumes equidistance between scale items, i.e. the distance between the first and second category is equivalent to the distance between the second and third. Data often do not meet this assumption because creating ordinal categories that consistently divide latent concepts is difficult. Second, this approach assumes that experts are exchangeable, i.e. they provide estimates of latent con- cepts that diverge only because of idiosyncratic differences in their perceptions of concepts.

A large body of literature contradicts this latter assumption.[3] Experts may perceive scales differently and thus report different values, even though they perceive the same level of a latent concept: my score of "one" may be your score of "two," though we both mean the same value. Equally importantly, some experts are likely more expert than others. For example, while the V–Dem Project employs a rigorous recruitment protocol designed to select coders with expertise regarding both the concepts and countries they code, it is inevitable that some experts will be less knowledgeable about a given concept or country than their peers. Given both of these concerns, aggregation algorithms of expert-coded data should account for differences in expert reliability, as well as differences in scale perception. Taking the mean over expert scores does neither.

In contrast, the V–Dem Project employs a Bayesian Ordinal Item-Response Theory (IRT) algorithm to aggregate expert coded data. While IRT models are commonplace in the social- scientific literature, their application in the expert-coded data context is more novel.[4] The V–Dem measurement model estimates a latent score for a given country-year-concept based on the ratings of the individual experts who code a case, i.e. it models each expert's coding as a reflection of the underlying concept. The model accounts for differences in scale perception by modeling experts as having idiosyncratic thresholds for placing a given concept on an ordinal scale (e.g. some experts are more apt to place an observation in middle or extreme categories than other experts). It also weights an expert's contribution to a latent concept estimate based on her level of agreement with other experts, conditional on her thresholds. This latter step rests on the assumption that experts are right on average, penalizing those experts who diverge from other experts, i.e. those who are less "reliable." While this assumption is strong, a lack of additional information by which to adjudicate between the opinions of dissenting experts necessitates this approach.

Figure 1 illustrates the utility of the V–Dem approach, using the Russian time-series (1900-2012) data from the variable "Freedom from Political Killings." Freedom from Po litical Killings is a five-point Likert scale question, with response categories ranging from "systematic political killings" (a value of zero) to "no political killings" (a value of four).[5] The subfigures present three different methods for summarizing these data. Subfigure (a) il- lustrates the raw coder values, with different colors representing the values provided by each of the nine Russia experts; the lines are a smoothed illustration of trends for individual ex- perts. This subfigure reveals that most experts are coding similar patterns: political killings in Russia were most systematic in the period 1924-1953, and were at their lowest level be- tween 1990 and 1999. Despite general agreement about trends in political killings over time, Subfigure (a) also indicates that

3. J. M. LeBreton and J. L. Senter. "Answers to 20 questions about interrater reliability and inter- rater agreement." Organizational Research Methods 11.4 (2007), pp. 815–852; Ren¨ e Lindst¨ adt, Sven-Oliver Proksch, and Jonathan B. Slapin. "When Experts Disagree: Response Aggregation and Its Consequences in Expert Surveys." 2016

4. J. M. LeBreton and J. L. Senter. "Answers to 20 questions about interrater reliability and inter- rater agreement." Organizational Research Methods 11.4 (2007), pp. 815–852; Ren¨ e Lindst¨ adt, Sven-Oliver Proksch, and Jonathan B. Slapin. "When Experts Disagree: Response Aggregation and Its Consequences in Expert Surveys." 2016

5. Graphics in this section reproduced from Kyle L Marquardt and Daniel Pemstein. "IRT Models for Expert-Coded Panel Data." Varieties of Democracy Institute Working Paper 41 (2017). They were created using data from V–Dem Dataset v6.2 and the R package ggplot2. Coppedge et al., V–Dem [Country- Year/Country-Date] Dataset v6.2 ; Hadley Wickham. ggplot2: elegant graphics for data analysis. New York: Springer, 2009.

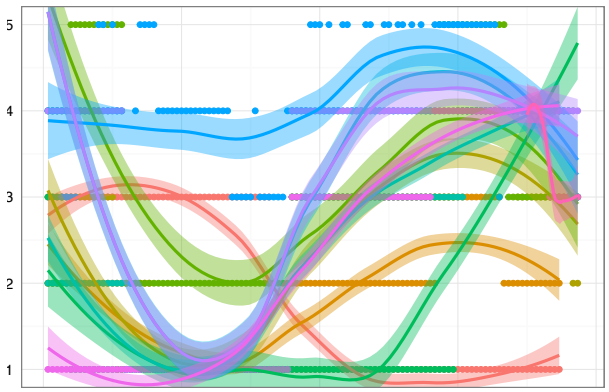
experts disagree about the scale: in most country-years, experts code nearly the full range of values. Finally, Subfigure (a) indicates that there is one expert who dissents from her peers, i.e. codes political killings as being least prevalent during Stalin's reign. This idiosyncratic trend is evidence that this expert is unreliable, misunderstood the scale, or is privy to secret information.

The problematic nature of using the mean and standard deviation to characterize such data is clear from Subfigure (b), which illustrates the raw mean and 95 percent confidence interval about this estimate over years. While the trends in the mean are consistent with historical events, the fact that the confidence intervals span the scale for most country-years presents problems for interpretation: strictly speaking, they indicate that levels of political killings in Russia do not differ significantly between any year across the past 115 years.

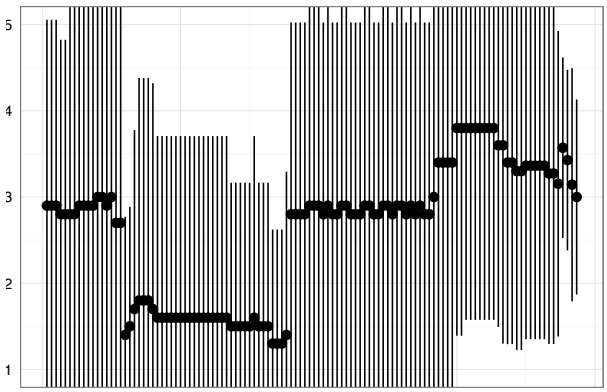
Subfigure (c) provides interval-level output from a modified version of the V–Dem model for the same data. In the subfigure, horizontal lines represent thresholds: scores below the lowest threshold indicate values in the category of one ("systematic political killings") and those above the highest threshold indicate values in the highest category ("no political killings"). Dots represent the median estimate for each year across iterations of an MCMC algorithm, while vertical lines represent the 95 percent highest posterior density interval, a Bayesian equivalent of confidence intervals. While output from this model show similar trends as the raw mean, by accounting for variation in scale perception and expert reliability, the model reports drastically lower levels of uncertainty and puts most of the posterior density for the Stalinist era below the lowest threshold, indicating high confidence that political killings were systematic during this time.

Simulation analyses have indicated that the findings from these descriptive analyses are generalizable: in simulations that involve variation in DIF and expert reliability, the V–Dem model outperforms both the raw mean and other IRT parameterizations in recovering the "true" latent concept values. Equally importantly, in

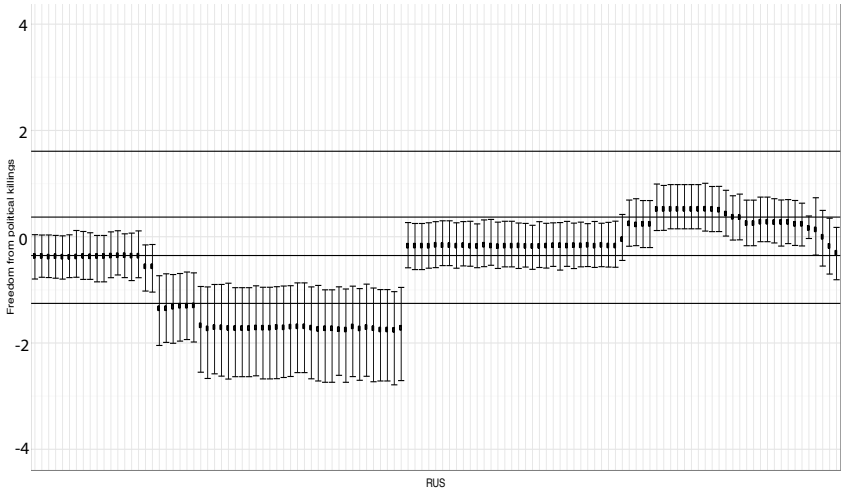
Figure 1: Freedom from political killings in Russia, 1900-2015



(a) Expert scores



(b) Raw expert mean



(c) V–Dem IRT model

simulations without these sources of variation, the V–Dem approach does not generally yield worse output than the raw mean.[6] In other words, while the V–Dem model may not be necessary in a world where experts perceive scales in the exact same way and do not differ in their level of reliability, it is a safe approach for the world in which we probably reside.

2 Measuring democracy with expert-coded data

This discussion has focused on low-level variables in the V–Dem dataset: variables measur- ing discrete concepts that are related to democracy, but which do not measure democracy writ large. However, the breadth of the V–Dem data presents the opportunity to aggregate these variables to create measures of higher-level concepts. Theory drives one method for doing so: a scholar both selects variables and derives aggregation rules based on her beliefs about a theoretical construct. A second method is inductive: algorithmically determining the relationship between a set of variables and an underlying factor of which they are all manifestations. Theory-driven approaches allow scholars to a priori determine the relationship between a set of variables and the construct they are measuring; the resulting measure will therefore have a clear interpretation and is unlikely to yield estimates that are a quirk of data structure. On the other hand, since theory rarely specifies precise mathematical formula for aggregation, these approaches require making difficult decisions about the relationship between manifest variables and the latent construct being measured.

6. Marquardt and Pemstein, "IRT Models for Expert-Coded Panel Data"

The V–Dem data present an ideal opportunity for analyzing differences between theory- driven and inductive approaches for aggregating data. Since almost all variables in the V–Dem dataset are related to concepts of "democracy," they represent a fertile ground for exploratory analysis. Project members have also developed a set of theoretically-derived measures of high level concepts—though there is no V–Dem measure of "democracy," V-Dem does provide five composite variables representing different forms of democracy.[7] At the basis of these five variables lies "Polyarchy," a measure that epitomizes the theory-driven approach to aggregating variables into a high-level concept.

2.1 Theory-driven approaches

Polyarchy is an aggregation of V–Dem data designed to proxy a Dahlian conceptualization of electoral democracy.[8] It includes five measures of institutional guarantees that Dahl enumerates as being essential to ensuring government responsiveness to its citizens. These measures include an index of the degree to which a government's executive is elected, the proportion of adult citizens eligible to vote, and output from three Bayesian factor analyses (BFAs) of expert-coded data. These BFAs measure the functioning of elections ("clean elections"), openness to civil society and polity party formation ("freedom of association") and freedom of expression. All of these variables are scaled such that they have values between zero and one; they are aggregated to Polyarchy by averaging the sum and product of the constituent variables. This multiplicative component of the equation represents the belief that the different measures are constitutive: a political system with a low score on one measure should receive a relatively low score, regardless of its other scores. The additive component of Polyarchy models the measures as being equally representative of an underlying concept, and thus deserving an equivalent degree of representation. By averaging the sum and product of the constituent elements, the index assigns equal weight to traditions that regard the constituent elements of Polyarchy as additive or constitutive.

The design of this index rests on well-developed theory regarding both the elements of polyarchy and how they fit together with regard to the concept. At the same time, the exact aggregation criteria requires additional assumptions. For example, assigning equal weight to the additive and multiplicative components of Polyarchy represents a reasonable adjudication between two theoretical traditions. However, other conceptual frameworks would assign greater weight to either component (the V–Dem dataset also includes the additive and multiplicative components of the index). Similarly, the determination of variables which enter into Polyarchy is restrictive: multiple civil society variables exist in the V–Dem dataset, but Polyarchy only incorporates two as an element in a larger composite index. A less restrictive approach

7. Michael Coppedge et al. "V–Dem: A New Way to Measure Democracy." Journal of Democracy 25.3 (2014), pp. 159–169; Michael Coppedge et al. "Measuring High Level Democratic Principles using the V–Dem Data." The Varieties of Democracy Institute Working Paper 6 (2015)

8. Jan Teorell et al. "Measuring Electoral Democracy with V–Dem Data: Introducing a New Polyarchy Index." The Varieties of Democracy Institute Working Paper 25 (2016); Robert A. Dahl. On Democracy. New Haven: Yale University Press, 1998

would potentially yield a more thorough measurement of the concept.

2.2 Exploratory approaches

At the opposite extreme of theory-driven approaches to aggregating democracy lies an en- tirely inductive approach. Here I present an example of such an approach, "Inductive Democ- racy."[9] Inductive Democracy is the output from a Bayesian latent variable analysis of all low-level V–Dem variables—both expert-coded and factual—that have high levels of time- series coverage (i.e. more than 50 percent of the country-years in the data set). While this approach has somewhat arbitrary inclusion criteria, it encompasses 180 variables, a much wider range than Polyarchy. The premise of this measure is that all these variables are reflections of some underlying concept, in this case "democracy;" the algorithm iteratively determines a unique interval-level variable that best explains variation across them. There are numerous flaws in this premise. For example, the input variables violate the assumption of local independence, which underlies latent variable analysis: for example, the presence of election management bodies likely increases the degree to which elections are free and fair, as opposed to both election management bodies and free and fair elections being only related insofar as they are reflections of the underlying concept of democracy. Equally importantly, some elements of democracy have greater representation in the dataset than others (e.g. there a fewer measures of direct democracy than civil society); the resulting measure may therefore overemphasize those variables relative to others since they correlate highly. Finally, some V–Dem variables have a tenuous relationship with many conceptions of democracy (e.g. measures of sovereignty). While variables less reflective of "democracy" should be less correlated with the other variables in the model that are more reflective of the concept—and thus receive less weight in the construction of the index—the fact that they are in the measurement procedure at all is a potential flaw in the approach.

2.3 Comparison of aggregation approaches

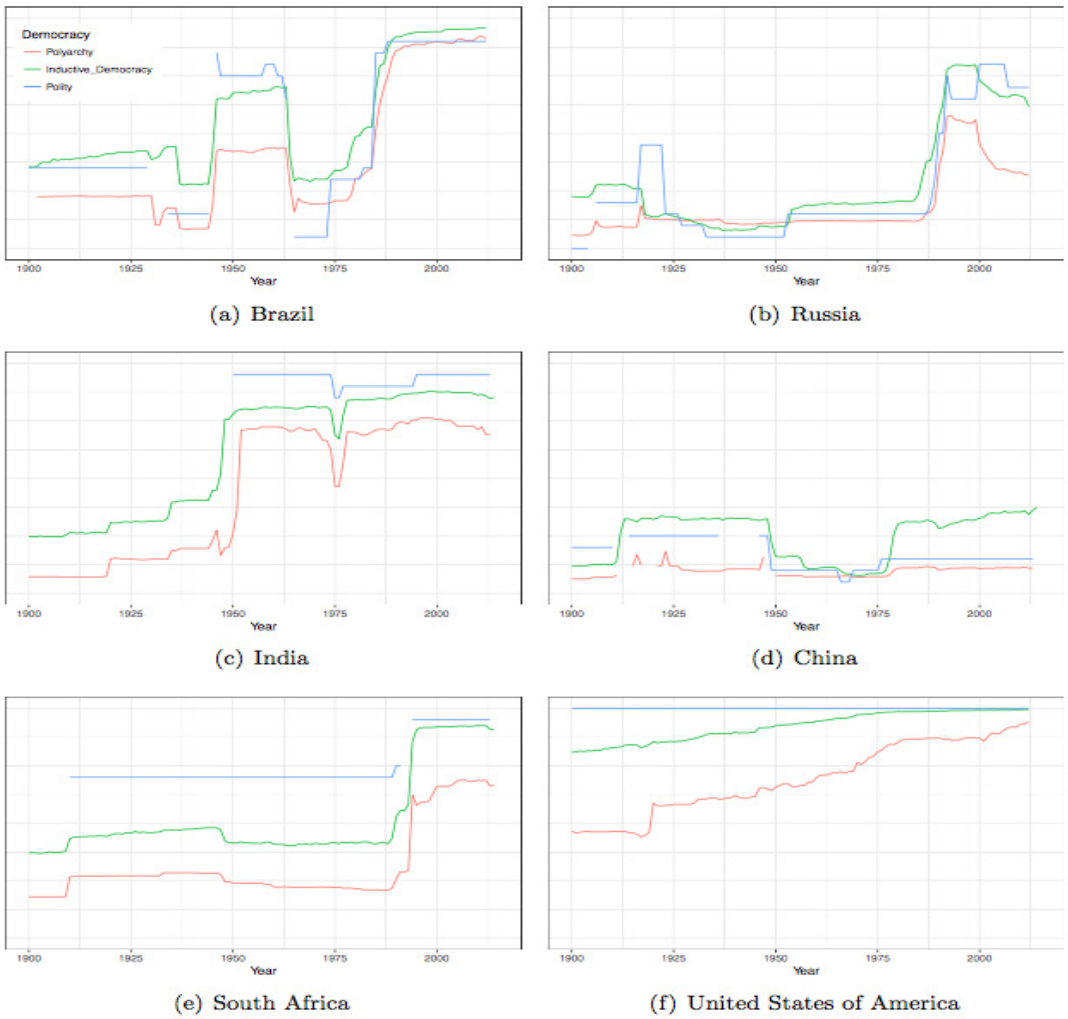
Polyarchy and Inductive Democracy are highly correlated: the Pearson correlation coefficient is 0.94. Such a high correlation indicates that the measures are measuring similar constructs; such correlations are in fact relatively common among measures of high level concepts such as democracy. Indeed, both measures are also highly correlated with Polity IV, a commonly- used measure of democratic governance: both Polyarchy and Inductive Democracy have a Pearson correlation coefficient of roughly 0.84 with Polity IV scores.[10] However, the high correlation belies important points of convergence and divergence, Figure 2 shows trends across time for all of these measures of democracy—as well as Polity IV scores, as a point of reference—in a diverse set of countries: Brazil, Russia, India, China, South Africa and the United States. I scale all measures such that they are between 0 and 1, using the CDF of the normal distribution for the inductive measures. Since the scales are not directly comparable, I focus on trends.

All three measures of democratic governance show similar trends, with substantive changes in levels occurring in parallel. However, two points of divergence are of note here. First, the fact that Inductive Democracy incorporates variables that measure different aspects of democracy than Polyarchy allows its scores to increase or decrease when the input variables for Polyarchy are constant or show less change. This trend is visible in cases like China after 1990 and Brazil from 1900-1929, where Polyarchy shows little change over time, while In- ductive Democracy shows a gradual-but-clear trend of increasing scores. Second, Polyarchy often evinces changes that are much more extreme than those in Inductive Democracy (e.g. in 1920 in the United States and in Russia in 1917). This difference is a function of Pol- yarchy's multiplicative structure: while a large number of variables must change a great deal for the level of Inductive Democracy to change, a substantial change in any of Polyarchy's constitutive variables

9. Input data from V–Dem Dataset v5. Michael Coppedge et al. V–Dem [Country-Year/Country-Date] Dataset v5. Varieties of Democracy (V–Dem) Project, 2015. Index created using R package runjags. Matthew J. Denwood. "runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS." Journal of Statistical Soft ware 71.9 (2016), pp. 1–25; Martyn Plummer. rjags: Bayesian graphical models using MCMC. 2012. Further details regarding the measurement procedure and a replication dataset are available at Harvard Dataverse: http://dx.doi.org/10.7910/DVN/JTLAIG.

10. Monty G. Marshall, Ted Robert Gurr, and Keith Jaggers. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2015. Center for Systemic Peace, 2016

Figure 2: Trends in different measures of democracy across countries, 1900-2015



(a) Brazil (b) Russia

(c) India (d) China

(e) South Africa (f) United States of America

results in a large shift in its value.

3 Conclusion

On one level, the similarities between Polyarchy and Inductive Democracy are heartening: the convergence of almost wholly inductive and theoretical approaches with regard to general trends indicates that, given high quality data and reasonable inclusion and aggregation criteria (be they inductive or theoretical), different aggregation approaches yield similar output. On another level, the fact that trends in certain key cases do diverge—as does the magnitude of change in values due to important historical events—-indicates that choice of aggregation criteria has important consequences that the high correlation belies. The correct aggregation criteria is therefore likely a function of the researcher's comfort with the theoretical assumptions—or lack thereof—employed in the construction of different indices.

ANDERSEN, CONTINUED
(continued from page 6)

blogs and on twitter with regards to the Trump presidency runs counter to the near-consensus in the scholarship on democratic breakdown and authoritarian regimes.

The resurgence of interest provides a welcome occasion to assess the scholarly consensus and evidence on the relationship between political culture and regime transition as well as to discuss how to measure political culture in a way that allows both cross-sectional and over-time comparison. To clarify, I understand political culture as deep-seated values and practices with relevance for the demands which citizens formulate in the political sphere, and which condition the nature and intensity of their political behavior. As such, electoral support for authoritarian or proto-authoritarian parties and candidates, protests against stolen elections, etc., also fall within the remit of political culture.

Re-evaluating the evidence on political culture and regime transition

One data point should not change political scientists' minds. If the evidence suggested political cultures predisposed against democracy could not lead to institutional change before Trump's inauguration on Jan. 20th 2017 (or his election on Nov. 8th 2016, when he accepted the Republican presidential nomination in July 2016, or when he first became the Republican primary frontrunner in the summer of 2015), then political scientists should not see it differently now.

So, even if fatigue and disillusion with liberal democracy are on the rise in Western countries and the rest of the world, do we need to worry?

In emerging and consolidated democracies, illiberal political forces can erode democracy or lead to its breakdown if they garner enough electoral support, as in Peru after 1990 and most spectacularly in Germany in 1933. Conversely, popular regimes – whether democratic or authoritarian – can wield support from their partisans as a form of coup-proofing when faced with coup attempts, as in Turkey's failed coup of June 2016.

Post-electoral protests in competitive authoritarian regimes often provide focal points for regime change if dissatisfaction is sufficiently widespread, and opposition actors can successfully coordinate anti-regime coalitions as in Serbia in 2000. Challenges by popular opposition candidates, dissidence and other acts reflecting a public lack of support for the regime are taken very seriously by authorities in all non-democracies. Even in closed one-party or non-electoral regimes, committed democrats can eventually subvert authoritarian regimes if they fail to legitimate themselves with strong existential narratives and effective propaganda as in much of Eastern Europe after 1989. [2]

Political culture and authoritarian strategies: repression, cooptation, performance and legitimation

Considering the set of fundamental strategies available to authoritarian regimes, mass support and political culture also appear to have strong implications for how regimes carry out (i) repression, (ii) cooptation, (iii) performance improvement, and (iv) legitimation.

Repression can be hard to undertake because security forces or armies sometimes refuse to fire against co-nationals who may also be co-religionists, co-ethnics or even family members. Finding ways around this culturally conditioned distaste for exercising repression is one the main tasks of authoritarian coercive machinery. In the British Empire, minority ethnic groups (such as Sikhs in India, Tamils in Sri Lanka and the Karen in Burma) were disproportionately recruited to the armed forces due to the greater ease with which they could be persuaded to use violence against rebellious, larger groups by colonial officials. Numerous examples of using armed forces from other regions or even from abroad exist also from Tunisia, Egypt, Bahrain and Syria in the Arab Spring and from the Maidan Revolution in Ukraine in 2014.

But even when regimes manage to create armed forces that use it, high-intensity coercion may also backfire and ignite yet more protest if it is broadly perceived

2. For consolidated and emerging democracies, see Juan Linz, The Breakdown of Democratic Regimes: Crisis, Breakdown and Reequilibriation (Baltimore: Johns Hopkins University Press, 1978); Aim Simpeng, Opposition to Democracy in Democratizing States: Thailand and Beyond, unpublished book manuscript. For post-electoral protests see Emily Beaulieu Electoral Protest and Democracy in the Developing World (New York: Cambridge University Press, 2014). For repression and collective action see Gary King, Jennifer Pan & Molly E. Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," American Political Science Review, 107(2) 2013: 326-343 and the authors' other articles on Chinese censorship and social media. For legitimation in closed regimes, see Martin K. Dimitrov, "Tracking Public Opinion under Authoritarianism: The Case of the Soviet Union during the Brezhnev Era," Russian History, 41 2014: 329-353 and Dictatorship and Information: Autocratic Regime Resilience in Communist Europe and China, unpublished book manuscript.

as illegitimate (such as in Ukraine in late 2013, Iran in 1978 or the 1960 Sharpeville Massacre in South Africa). On the other hand, law and order policies dispensing repression, especially against outgroups, can be surprisingly popular in both democracies and non-democracies.[3]

Cooptation of elites – whether in opposition parties, business, trade unions, religious organizations, or regionally – can pre-empt public criticism, opposition coordination, or social mobilization and thus keep authoritarian regimes in power. Yet, some actors or groups are harder to coopt than others, for instance due to strong ideological commitments or to notions of cultural belonging and difference as with nationalist, secessionist or regionalist movements. Clientelism is also a type of citizen cooptation by embedding them in long-term, transactional networks, whereby they are induced by the provision of gifts, access to public services or even public employment to support the regime. As at the elite level, however, "good democrats make bad targets" for cooptation or clientelism, and conversely citizens often demand clientelist politics themselves.[4]

Performance improvement is a ubiquitous survival strategy for authoritarian regimes. While often highly effective, improving policy performance is not a panacea: in some societies, citizens care more about material outcomes than in others, where the processual aspects of decision-making sometimes take primacy regardless of the policy results produced by the regime. This point has most famously been made by Ronald Inglehart in his work on post-materialism.

Finally, legitimation is the process of seeking legitimacy from the public by communicating and arguing for the worth and appropriateness of the government and regime more broadly. Unlike repression, cooptation, and performance, then, regime legitimation is entirely and not only partially about culture and popular support. On the one hand this includes propaganda, information distortion and media control, while on the other are various substantive types of non-democratic legitimation that rely heavily on highly context-specific repertoires of historical and culturally salient references. Thus, regimes seek firstly to manipulate information about its performance that might lead citizens to evaluate them unfavorably given preexisting norms of legitimate government, and secondly to alter these norms in a direction that is more manageable for the existing regime, and wherein it has a comparative advantage.

Non-democratic regimes seek to legitimate their rule based on many pillars. Procedurally, they may claim legitimacy based on (i) personality cults of charismatic leaders, (ii) ideological precepts, (iii) traditionalist appeals to history in long-lived regimes, (iv) bureaucratic norms of technocracy and expert rule, or even (v) religious principles for leader selection. Substantively, a non-exhaustive list of the most commonly claimed bases of legitimacy includes (i) law and order in post-conflict, high-crime or ethnically tense contexts, (ii) economic performance, (iii) economic populism in highly unequal settings, (iv) nationalism, often directed against foreign threats, or (v) ethnic or ethno-religious identity against internal threats.[5]

As I have shown, political culture is relatively easily integrated into existing analysis of regime change processes from democratic breakdown to democratic transition in all types of regimes and into analysis of the strategies undertaken by authoritarian regimes to stay in power. If we think these strategies matter in and of themselves, then political culture should as well.

How to measure political culture?

How should we then measure political culture in a way that is relevant to comparative democratization as I have described it in the two preceding sections?

I divide measures of political culture into those that are (1) textual, or expressive of sentiment and preferences, and those that are (2) behavioral and measure action reflecting a given political culture. This applies for all types of methodologies and data types: qualitative and quantitative, and

3. Subhasish Ray, "The Nonmartial Origins of the 'Martial Races': Ethnicity and Military Service in Ex-British Colonies," Armed Forces & Society, 39(3) 2012: 560-575; Brian Martin (2007). Justice Ignited: The Dynamics of Backfire (Lanham, MD: Rowman & Littlefield, 2007); Jane Esberg, "The Audience of Repression: Killings and Disappearances in Pinochet's Chile," working paper (2016).

4. Steven Levitsky & Lucan Way, "The Durability of Revolutionary Regimes," Journal of Democracy, 24(3) 2013: 5-17; Ryan E. Carlin & Mason Moseley, "Good Democrats, Bad Targets: Democratic Values and Clientelistic Vote Buying," Journal of Politics 77 (1) 2015: 14-26; Simeon Nichter & Michael Peress, "Request Fulfilling: When Citizens Demand Clientelist Benefits," Comparative Political Studies, OnlineFirst.

5. Johannes Gerschewski, "The Three Pillars of Stability: Legitimation, Repression, and Co-Optation in Autocratic Regimes," Democratization, 20(1) 2013: 13-38; Max Weber, Economy and Society: An Outline of Interpretive Sociology (Berkeley, CA: University of California Press, 1922 [1978]); David Easton, A Systems Analysis of Political Life (New York: Wiley, 1965); David Beetham, The Legitimation of Power (London: Palgrave Macmillan, 1991); Bruce Gilley, The Right to Rule: How States Win and Lose Legitimacy (Princeton: Princeton University Press, 2009); Christian V. Soest & Julia Grauvogel, "How Do Non-Democratic Regimes Claim Legitimacy? Comparative Insights from Post-Soviet Countries," GIGA working paper (2015).

for the latter both in experimental and observational approaches.

For quantitative observational and qualitative studies (whether interpretivist or seeking causal inference), it is important to be able to measure aspects of political culture reliably and validly between units (whether at the individual, subnational or national level) and over time. For lab experiments, comparison is almost always between individuals, and measurements are usually very close in time. For field experiments or randomized controlled trials (RCTs), researchers are sometimes interested in long-term effects of interventions, and thus precise measurement over time of differences between groups, cities or other units is also important.

(1) Political culture measured using texts

The quintessential measure of political culture in quantitative studies are survey data. Tellingly, Michael Coppedge's 2012 monograph Democratization and Research Methods includes a 35-page chapter entitled "Political culture and survey research," which only in passing makes reference to other ways of studying political culture.

Comparative, cross-national survey data were pioneered by Ronald Inglehart and the World Values Survey and later supplemented by regionally specific survey data collection efforts like the Afrobarometer, the European Social Surveys as well as others. With adequate sampling, surveys are representative, and allow for relatively easy comparison across individuals, groups, administrative units, countries, etc. Downsides are that they survey data are relatively superficial and plagued by various response biases – as well as their high financial cost. They are also somewhat artificial and unrepresentative of real life politics and culture, which develop in groups and highly emotionally charged contexts.

For a long time, survey data were one of the only types of comparative, quantitative data available for the study of political culture. However, large-scale analysis of texts such as parliamentary speeches, media reports, books or social media posts has recently become feasible. Especially automated text analysis holds a treasure trove of possibilities for the study of texts, which are primary bearers and expressions of political culture.

Supervised methods of automated text analysis usually compare a corpus of texts to a reference text and score their similarity to it such as in Richard Nielsen's work on Jihadi radicalization of imams, whether this is in terms of a continuous or categorical variable. Unsupervised methods instead identify textual structures and similar passages without a set of hand-coded texts to place texts on a scale or assign subsets of text to mutually exclusive categories. In this way Lisa Blaydes, Alison McQueen and Justin Grimmer have examined the 'mirrors for princes' genre in the Christian and Islamic worlds, and when in time certain themes rose or fell in importance in elite political culture. Others study mass political culture in similar ways using social media posts.

In qualitative studies of political culture, texts such as speeches or information gleaned from interviews with elite figures or regular people have been the dominant measures, as in Frederick Schaffer's work on meanings of democracy in Senegal.[6] Today, political communication often either takes place online or is accessible online, increasing the reach and ease of qualitative analysis of political culture.

(2) Political culture measured using behavior

Behavioral measures of political culture have become increasingly popular, especially in quantitative research. Economists have long been skeptical of stated preferences, preferring 'revealed preferences' such as those in making a purchase or sharing resources in experiments like the well-known dictator, ultimatum or trust games. Qualitative researchers have also long studied costly signals in political culture such as personality cults, propaganda, and architecture, political branding and other symbols, and how they are perceived by elites and masses. Lisa Wedeen's work on Syria and the preceding Hafiz al-Assad regime is an example of such work focusing not just on textual sources, but also on behavioral, non-textual measures of political culture.

Quantitative, observational research has begun to develop new behavioral measures, too. Studies

6. Richard Nielsen, Deadly Clerics: Blocked Ambition and the Paths to Jihad (New York: Cambridge University Press, forthcoming); Lisa Blaydes, Allison McQueen & Justin Grimmer, "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds," working paper (2016); Amaney Jamal et al., "Anti-Americanism and Anti-Interventionism in Arabic Twitter Discourses," Perspectives on Politics 13 (1) 2015: 55–73; Frederick Schaffer, Democracy in Translation: Understanding Politics in an Unfamiliar Culture (Ithaca: Cornell University Press, 2000).

have examined names for the ethnic and other identities they betray, buildings with a cultural significance such as churches, and google searches that display particular cultural attitudes or biases. As an example, Nico Voigtländer & Joachim Voth use medieval pogroms and anti-Semitic Judensau sculptures at the city-level as measures of anti-Semitic cultural traits in the past and attacks on synagogues as well as letters to the editor sent to the anti-Semitic magazine Der Stürmer to measure it during the Nazi era.[7]

The future of the empirical study of political culture will likely involve lots of behavioral measures associated with automated, digital data collection (such as google searches, online purchases, open government data, etc., often with precise geocodes) and the development and accumulation of large datasets.

Political culture after Trump

In an era of decreasing popularity of liberal democracy, clarifying what its declining legitimacy means for democracy worldwide – and how to measure it – is a crucial task for political scientists and particularly for scholars of comparative democratization.

I have suggested several strategies to integrate political culture into the analysis of regime change processes by focusing on electoral support, mass party coup-proofing, post-electoral protests and de-legitimation of authoritarian narratives. I have also shown how aspects of political culture are relevant for repression, cooptation, policy performance and legitimation, four widely studied strategies of authoritarian rule.

Further, I have given an overview of how political culture may be measured today using either textual or behavioral data sources and how automated, digital data collection and automated text analysis will likely change how we measure political culture in political science.

After a long hegemony of democratic ideas, the coming years will likely provide us with lots of data to test how resilient new and consolidated democracies really are to major shifts in political culture.

7. Lisa Wedeen, Ambiguities of Domination: Politics, Rhetoric and Symbols in Contemporary Syria (Chicago: University of Chicago Press, 1999); Nico Voigtländer and Joachim Voth, "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany," The Quarterly Journal of Economics 127 (3) 2012: 1339-1392.

Carlitz, Harris, Kao, Landry, Lust and Rakner, CONTINUED

(continued from page 7)

research also tends to overlook the fact that most developing countries function as dual polities where community-based and traditional institutions coexist and interact with state institutions. The governance literature has either ignored such dualism or perceived of it as a problem. As a result, few scholars have examined the overlapping functions and interactions between state and non-state institutions, particularly at the local level.

This paper proposes a new way to conceptualize and measure governance that addresses the limitations of previous approaches.[1] We begin by reviewing the literature to illustrate dominant approaches to the definition and measurement of governance. After highlighting some of the short- comings with the conceptualization and methodology of extant approaches, we present a more holistic way of thinking about governance and propose an alternative measurement strategy. We illustrate the utility of our new measure with the case of Malawi.

2 Dominant Approaches to Defining and Measuring Governance

As noted above, the observation that countries with similar policies (on paper) often exhibit vastly different development trajectories has driven the recent surge of interest in "good governance" among both foreign aid donors and academics. The first generation of governance metrics therefore aimed at identifying the distinct dimensions of government performance that were most conducive to economic development.[2] Closely related to the notion that good governance begets economic growth is the idea that foreign aid will be more effective in better governed countries. Reflecting this, the World Bank since 1998 has included governance criteria in its Country Policy and Institutional Assessment (CPIA) framework, which informs the allocation of the Bank's concessional International Development Association (IDA) loans. Given political sensitivities, CPIA data are not publicly available for academic research. However, in the 1990s the Bank supplemented these internal measures with the publicly available Worldwide Governance Indicators (WGI). Since 1996, the WGI have annually assessed the quality of governance in over 200 countries based on a combination of expert survey data and objective indicators.

Defining good governance in terms of the ability of governance institutions to promote eco- nomic growth (or aid effectiveness) borders on tautology, however, as one cannot define governance in such a manner without first measuring its effects. In order to avoid such circular logic, Rothstein and Teorell define quality of government (QoG) as the "impartiality of institutions that exercise government authority."[3] Such impartiality relates to both the "input" side (access to public au- thority) and the "output" side (the way in which authority is exercised). The authors explain that impartiality on the input side implies political equality, which is a basic norm for legitimatizing democracy. However, they argue that democracy is a necessary but insufficient criterion of QoG, since there is no guarantee that majoritarian rule will respect the impartiality principle when it comes to how government power is exercised. Impartiality on the "output side" rules out all forms of corruption as well as clientelism, patronage, nepotism, political favoritism, and discrimination. It also implies and entails the rule of law.

The QoG Institute at the University of Gothenburg has compiled an extensive dataset that includes a range of variables capturing both the core features of QoG (impartiality, bureaucratic quality and corruption) as well as broader measures (rule of law and transparency). The indicators include various measures of corruption, bribe-paying, performance of democratic institutions, and state legitimacy. They are based primarily on expert surveys and focus largely on state institutions, covering around 200 countries for every year from 1946 to the present day.

2.1 Critiques of Dominant Approaches

The governance metrics reviewed above have engendered a fruitful array of scholarship, including the development and testing of theories about the determinants and consequences of good governance. However, they sama amauffer from a number of limitations – both in terms of the methodologies employed and the broader conceptualization of goverance – which have potentially serious implications for the practice of development.

2.1.1 Methodological Issues

A number of scholars have argued that extant governance indicators are methodologically flawed given their reliance on expert surveys, which can contain systematic biases. For example, external experts tend to be more impressed by announcements of reforms than by their actual implemen tation.[4] In addition, experts

1. This paper draws on a longer, piece by Ruth Carlitz, Kristen Kao, Pierre Landry, Ellen Lust, and Lise Rakner, entitled, "Unpacking 'Good Governance': Exploring Subnational and Cross-Sectoral Variation in the Quality of Governance in Malawi," and will be presented at the 2017 Annual Conference of the European Political Science Association.

2. Stephen Knack and Philip Keefer. "Institutions and economic performance: cross-country tests using alternative institutional measures". In: Economics & Politics 7.3 (1995), pp. 207–227; Rafael La Porta et al. "The quality of government". In: Journal of Law, Economics, and organization 15.1 (1999), pp. 222–279.

3. Bo Rothstein and Jan Teorell. "What is quality of government? A theory of impartial government institutions". In: Governance 21.2 (2008), pp. 165–190, p. 165.

4. Matt Andrews. The limits of institutional reform in development: Changing rules for realistic solutions. Cam- bridge University Press, 2013.

have been shown to systematically overstate the level of corruption experienced by citizens and to rank countries according to their own political preferences.[5]
The WGI have also been criticized in relation to their comparability over time. They are rescaled to have the same global averages in every period, although the WGI's creators argue that there is little evidence of significant trends in world averages of governance, making such rescaling inconsequential. Concerns have also been raised about the independence of the different data sources underlying the WGI and the consequences for the aggregate indicators.[6]

2.1.2 Conceptual Issues

Beyond methodological critiques, many widely cited governance indicators raise questions in terms of the main elements of their definition – for example, including democracy as a feature of good governance. As above, this risks setting up a tautology as it does not allow us to test the assumption that democracies are in fact better governed than other types of polities.

Furthermore, the dominant approaches to defining and measuring governance, and policy pre- scriptions emanating from them, tend to reflect a one-size-fits-all model of government. Even if we agree that building formal, rule-based systems of democratic governance is a valid long-term goal, expecting low-income countries to "catch up" with better-governed Western nations that have experienced very different histories may be unrealistic. Furthermore, using development assistance to promote "good governance" as defined above may fail to provide poor countries with effective strategies for improving security, encouraging investment, or delivering public services. Hence, it may be more useful to begin by looking at the desired outcomes of good governance, and then work backward to see what institutional arrangements are associated with better results. The approach to measuring governance proposed in this paper facilitates such a strategy.

Moreover, the central state need not be the primary unit of analysis. Given that almost every country in the world has by now implemented some kind of decentralization reform, examining variation in governance quality across local government units represents an important line of inquiry, and one that has largely been unexplored.

Relatedly, the governance literature tends not to take the multiplicity of governing structures in many developing nations into account. As Lust and Ndegwa explain, "actors are engaged variously in cooperation and competition with each other in the struggles over power and resources located in various arenas – both within and outside the state, and individuals are subjected to – and trying to benefit from – multiple authorities and institutional structures.[7]Relatedly, there is an emerging literature on traditional leadership suggesting that traditional leaders and institutions (i.e. tribal chiefdoms, religious leaders and their followers, etc.) may play a critical role in brokering local development projects and that traditional leaders may have a greater capacity to organize responses to rural problems than elected politicians.

Finally, the rules and institutions that shape governance and outcomes may vary across different "imagined" communities (defined in terms of shared social identity such as ethnicity or gender). Ultimately, it is important to recognize that behaviors are affected – and ultimately community outcomes are driven – by various sets of rules and norms. The extent to which such rules and norms diverge is an interesting empirical question that is likely to vary over time and space. In focusing on one dimension or set of institutions shaping governance at the national level, the dominant approaches described above miss important factors that lead to both subnational variation and differences across countries with otherwise very similar political institutions.

3 Suggestion for a New Approach

Responding to the critiques enumerated above, we propose a new way to conceptualize and measure governance. This approach can help us understand why some communities, in similar national contexts, succeed in delivering quality public goods, while others are not able to provide the most rudimentary services – a question the dominant strategies for defining and measuring governance cannot answer.

3.1 Conceptualizing the community

We begin by thinking seriously about the different communities over and within which we expect governance to vary. As noted above, subnational government and administrative jurisdictions represent an obvious starting point. In addition, we might expect different identity-based groups to govern themselves in different ways. Hence, it is important to consider non-geographically bound "imagined" communities.

One basis for imagined community that has received considerable attention in the political science and economics literature is ethnic identity.

5. Mireille Razafindrakoto and Fran¨ cois Roubaud. "Are international databases on corruption reliable? A compar- ison of expert opinion surveys and household surveys in sub-Saharan Africa". In: World development 38.8 (2010), pp. 1057–1069.

6. Christiane Arndt and Charles Oman. "Uses and Abuses of Governance Indicators". OECD Development Centre Studies. 2006; Stephen Knack. "Measuring corruption in Eastern Europe and Central Asia: a critique of the cross- country indicators". World Bank Policy Research Working Paper 3968. 2006.

7. Ellen Lust and Stephen Ndegwa. "Layered Authorities: Beyond State-Centric Theory and Develpment Policy". Presented at The State in the Middle East, 100 Years After Sykes-Picot, American University in Cairo, March 25-26, 2017. 2017, pp. 2-3.

Scholars have focused in particular on the coordinating role of ethnicity in explaining voting behavior and the distribution of political benefits as well as the negative impact of ethnic diversity on public goods provision. An implied if not explicitly acknowledged assumption of most of these studies is that the rules governing co-ethnic and non co-ethnic behavior are the same across ethnic groups in a given context. For example, studies of ethnic favoritism in sub-Saharan Africa tend to treat patterns of favoritism as equivalent across leaders of different ethnicities. Differences in the motives driving such favoritism, or the technologies employed, are not explored. This reflects a general pattern in this literature, where the differences across ethnic groups (or even differences in norms/practices across more broadly defined groups) are largely overlooked.

However, differences across ethnic groups – and variations in the rules and norms that shape their social relations – can have important implications for governance and public goods provision. For example, different practices in lineage structures (e.g., matrlineality and patrilineality) are found to have implications for birth spacing, land investment, and economic development. Similarly, marriage practices, including the choice of partner (e.g., variations on endogamy and exogamy) and locality of the household (e.g., matrilocality and patrilocality) can impact investment and conflict. There is also reason to believe that people from the same ethnic group may adopt different social institutions in different places.

Different religious traditions have also been shown to have implications for governance. This line of thinking dates back to Max Weber's notion that Protestantism encouraged a particular mind- set that made one's occupation a calling from God, thus legitimating the acquisition of wealth. More recently, scholars have conducted empirical, cross-country tests of Weber's theory, demon strating, for example, that corruption is lower in countries whose largest religious group in 1900 was composed of Protestants.[8] Religious affiliation can also influence governance through patterns of discrimination that affect the construction of different religious identities.

Finally, gender represents another important basis for the formation of imagined community, with implications for governance and public goods provision. Perceptions of corruption tend to be lower in countries where women hold a larger share of parliamentary seats.[9] Scholars have also shown that women tend to engage in or condone bribery at lower rates. That said, such differences are not universal. For instance, women are less tolerant of corruption than men in Australia, while equivalent gender differences have not been found in India, Indonesia, and Singapore.[10]

3.2 Components of Governance

Once the communities of interest have been defined, we can think about governance in terms of two component parts: First, the nature of governance (NOG) refers to the power structures and rules for solving collective action problems and providing collective benefits to the community in question. These include, for example, modes of leader selection, opportunities for public participation, and reliance on personal networks.

Next, the legitimacy of governance (LOG) encompasses community members' perceived legit- imacy of and satisfaction with different modes of governance. For instance, if you have to pay to receive a particular service, is this understood as legitimate or a bribe? Satisfaction is tied to outcomes but also expectations, which are somewhat endogenously determined. That is, where bribes become commonplace, they may become to be seen as 'normal,' and similarly, when there is no transparency, people may come to expect this and even be satisfied. Thinking about governance in this way allows us to capture the subjective experience of governance from the community itself, in stark contrast to existing measures of governance.

Note that the first element of this definition is non-normative. The second element is normative to the extent that it is based on the norms of the community in question itself. This helps us to escape the trap of imposing a linear model based on the experience of Western countries and more organically understand differences in governance across communities. This also facilitates a more inductive means of identifying the features of governance arrangements most likely to lead to "good" outcomes. That is, one can first examine which communities exhibit the highest levels and most equitable degrees of service provision and then work backwards to identify the institutional features (state and social) that are associated with better outcomes. For instance, are these communities characterized by greater opportunities for public participation? Are people more likely to rely on state or non-state channels to solve their problems?

4 Application

As a first take at empirically testing the governance framework we develop above, we examine how governance varies with geography in Malawi,

8. Charles M North, Wafa Hakim Orman, and Carl R Gwin. "Religion, corruption, and the rule of law". In: Journal of Money, Credit and Banking 45.5 (2013), pp. 757–779.

9. David Dollar, Raymond Fisman, and Roberta Gatti. "Are women really the "fairer" sex? Corruption and women in government". In: Journal of Economic Behavior & Organization 46.4 (2001), pp. 423–429.

10. Vivi Alatas et al. "Gender, culture, and corruption: Insights from an experimental analysis". In: Southern Economic Journal (2009), pp. 663–680.

as well as by ethnicity, religion, and gender. In order to do this, we draw on the Local Governance Performance Index (LGPI) survey implemented in 2016 by the Program on Governance and Local Development (GLD) at the University of Gothenburg. The LGPI is a heavily clustered survey that allows us to uncover important local-level variation in governance and service provision. This sample design ensures a representative sample of the population at the local level. It differs from most nationally representative surveys that capture just a handful of respondents per municipality, prohibiting us from drawing conclusions about community drivers of governance outcomes. The LGPI uses household surveys that gather micro- level data from multiple communities, including data on experience, perception and satisfaction regarding cross-cutting governance issues. Of particular note is the LGPI's focus on experience rather than simply perceptions, allowing for an in-depth assessment of institutional quality and capacity, and a detailed mapping of institutional weaknesses and strengths from the perspective of ordinary citizens.

We explore subnational variation in the nature of governance by examining respondents' answers to the following two questions:

1. Would you say that it is very likely, likely, neither, unlikely or very unlikely that members of this village/ neighborhood could be counted on to "do something" if children were being disrespectful to other people or property?

2. Do candidates in your community pass out gifts, money, or promise access to government services during election time?

These questions were selected since they were asked of nearly all 8,000 respondents in the survey, and exhibit considerable variation. Given that some of these variables rely on different scales, they have all been recoded to range from -3 to +3.

We find that women are significantly less likely to say they would discipline others' children, although the effect of gender is not substantively large. We do not find any impact of ethnicity on this outcome, although we do find some variation by religious group. Specifically, people who identify as Christian or with a traditional/ ethnic religion are more likely to say they would discipline other people's children, than people identifying with other religious groups. However, all of these effects are dwarfed by village and district fixed effects.

Moving on to reported vote-buying, we find that women and older people are slightly less likely to report vote-buying, while people who identify as Muslim and respondents who identify as Tumbuka are more likely though the effect is substantively small. Again, however, these are dwarfed by the district and village effects.

We find similar results when it comes to questions regarding the legitimacy of governance (e.g., the extent to which people trust various authority figures or think they act fairly, perceived secrecy of the vote). Over all, we find highly geographically clustered and locale-specific variability in both the nature of and perceived legitimacy of governance. Notably, we find very little evidence for the impact of ethnicity on governance in and of itself. This is somewhat surprising given the focus of the African politics literature on the importance of ethnicity. In future work we intend to examine how the ethnic composition of different villages affects governance, as well as examine the impact of heterogeneity in terms of other identity groups. We will also consider how differences in the nature and legitimacy of governance relate to service delivery, in order to inductively identify the features of governance most likely to be associated with "good" outcomes.

social media usage that are presumed to coincide with a healthy democratic society.

2 Measuring Democracy Directly with Social Media Data

In a separate project, Jonathan Ladd of Georgetown University and I have been examining existing datasets for measuring the strength of democracy cross-nationally. This has involved a lot of the usual suspects – e.g., Freedom House, Polity, the Economist Intelligence Unit – as well as some important newcomers, e.g., V-Dem, the Unified Democracy Scores, and the Democracy Barometer. At the end of the day, however, almost all of these involve trying to assign a yearly score for each country in terms of how "democratic" the country is at that point in time, either in an overall "democratic" sense or in a more disaggregated component of democracy. A variety of tools are used to arrive at this score, but they largely involve relying on some form of expert analysis across more or less well defined categories that may or may not be aggregated across multiple experts. So the question is, could we somehow use social media data to replace the expert analysis, and go straight from the posts of individual politicians and/or citizens to a measure of "democracy"?

The answer is possibly, but it would certainly be a non-trivial task with serious methodological challenges. That being said, the potential payoffs from both a scholarly and policy perspective from pulling this off would also be non-trivial.

To come up with a single "how democratic is society" measure from social media would most likely require a three-step process. First, we would need to begin with pre-existing human coded measures of democracy. Next, we would need some collection of social media from each of these countries, say a random sample of tweets from the population over the period of time in questions (here, years). The final step would be to see if we could train a machine learning model to essentially to predict the level of democracy from the social media data.

Such a method would be fraught with difficulties. The data collection (and storage) for such a project would likely be enormous if we chose to rely on data from the mass populace.[3] This should not be under-emphasized: the cost and data storage challenges associated with working with something like the Twitter Firehose (i.e., the entire collection of Tweets being collected in real time) would make such a project reliant on serious financial backing. Another major challenge would be that we would need to train the model using text in many, many different languages, which might push scholars to rely more on metadata (which is language invariant), such as the timing of social media posts, or perhaps the geographic diffusion of social media posting. And all this precedes the point that we have no idea whether there are indeed enough "clues" in social media data to distinguish posts in democratic countries from posts in non-democratic countries. So there would need to be a huge investment in both time and resources before we had any idea of the ultimate viability of the project. Finally, to the extent that we were ultimately interested simply in the kinds of measures we've had previously (how democratic is country X in year Y), then this massive investment in infrastructure and capabilities would simply be to replicate other ongoing studies.

All that being said, however, there are two potential payoffs worth noting here. The first is that if – despite all the challenges identified in the preceding paragraph – we could indeed train a model to produce, for example, a measure equivalent to a Polity (or Freedom House, or even Unified Democracy) Score from social media data, then once that system was in place, there would be an opportunity going forward for enormous cost savings in terms of producing these measures in the future. So if, hypothetically speaking, we were entering a world where funding for Polity and Freedom House were likely to disappear, social media generated democracy scores would be one way we could continue to have these measures going forward.

The second potential advantage is that once we are using social media to generate democracy scores, we would no longer be limited to yearly measures. In fact, we could begin to disaggregate our measures at any interval in which we were interested. Further, we would no longer need to wait for ex-post analysis by experts to get our democracy scores; they could, in theory, be produced on a daily basis based on the previous day's data. Combining these two points yields the possibility of the perhaps the most tantalizing aspect of moving to a social media based approach to measuring democracy, which would be some sort of early warning system for countries in which

3. One way out of this conundrum would be to train a model based not on social media usage by the population but rather by elites, such as members of the legislature and government ministers. While still a very serious undertaking – finding the social media accounts of members of the government from a single county is a time consuming process, and remember these people (and accounts) are always changing – the data storage and retrieval demands would be potentially more manageable.

the baseline level of "democraticness" was changing.

However, these potential payoffs also raise a key final caveat of which any effort in this regard must be cognizant, which is that social media itself is dynamic. Training a model on data from 2015 does not guarantee that that model will continue to be relevant in 2017 not so much because the machine learning technology itself might be flawed, but rather because the very signals the model might have learned to pick up in 2015 to identify a more democratic society may no longer be present in the data in 2017. Consider, for example, a world in which emoji are replaced by stickers; a model trained on emojis might simply cease to be relevant. And herein lies the rub: if the benefit of a social media based approach to measuring democracy is that it could replace the very measures that were used to generate the model in the first place (as in the loss of funding scenario above) or so that it could produce democracy estimates in real time but that the model constantly needed new human coded data to make sure it was still functioning as intended, how valuable would these new contributions turn out to be in practice?

However, even if using social media data to simply produce a "democracy score" proves untenable, there are still other opportunities for using these data to aid in our measurement of democracy more indirectly (or, perhaps more accurately, in terms of producing components of a democracy measure). I briefly address three such possibilities in the remainder of this essay.

3 Using Social Media to Measure Attitudes Towards Democracy

The first, and perhaps most obvious, opportunity is to use social media data to measure societal support for democracy. While this is not normally considered a measure of democracy in and of itself, this is certainly a component of at least some definitions of democracy.

While minimizing the demands we are placing on social media data as compared to the taller goal considered in the previous section, using social media data to measure public opinion is no small task either. This is a topic we have taken up in much greater detail in a chapter in the forthcoming Oxford Handbook of Public Opinion[4]; in the remainder of this section I briefly summarize a few of the relevant points made there.

The challenges of using social media data to measure public opinion are by no means trivial. In our chapter, we highlight three in particular. First, there is the difficulty of measuring "opinion" without the structured survey questions to which we have all grown so accustomed. Thus if, for example, we wanted to measure "support for democracy", this would have be a two step process: find relevant discussions of democracy; and then run some sort of sentiment analysis that classifiers whether the discussion is positive or negative. Both tasks are challenging, especially if we are trying to disentangle "support for democracy generally" from "support for the performance of the current government" in a democratic regime. [5]

Second, even if we are able to solve the initial challenges of extracting positive or negative attitudes towards democracy from the text of social media posts, we next have to wrestle with the fact that social media users are not a representative sample of any population.[6] Thus any attempt to project from "attitudes on social media" to attitudes in the population at large will need to perform some sort of adjustment for non-representativeness. Finally, there other platform specific issues related to aggregating from individual to collective measures of public opinion, including what is made public and what is kept private, as well as limits on data collection. [7]

At the same time, the promise of social media as a means of measuring public opinion are similar to those for measuring democracy: real time data available from around the world and which can be disaggregated to just about any time period. And given the costs of conducting representative surveys cross-nationally, the cost-savings here may be even more significant than those discussed in the previous section.

4 Using Social Media to Measure Indicators Associated with Democracy

Using social media to measure public opinion is challenging. However, it may be possible to use social media data to measure other indicators that are generally associated with democratic regimes, but which don't have the demands as trying to measure opinions.

One promising candidate could be something like the ratio of tolerant to intolerant posts in a random

4. Klašnja et al., forthcoming.

5. See Pop-Eleches and Tucker (2017) for more on this distinction.

6. Occasionally, though, they do represent an entire population of interest, such as in studies of members of the US Congress, where today all members have a social media presence.

7. See Klašnja et al. (forthcoming), p.10-11 for more detail.

subsample of social media posts. Such a measure would not face the challenge of trying to identify "opinions towards democracy", but could perhaps serve as an important indicator on its own for the strength of democratic values in a country. We actually have an in progress paper in our lab (the NYU Social Media and Political Participation (SMaPP) lab) that utilizes exactly such a measure, albeit in a single country. [8]

On a grander scale, one could imagine companies such as Facebook publishing indices based on internal data that might not be available to scholars at the individual level for privacy reasons, but which could be released for public policy/academic outreach efforts as aggregate level indicators. Such measures could play a role in developing next generation democracy indicators.

5 Social Media Usage Patterns Associated with Democracy

The final "indirect" way we might want to think about using social media to measure democracy would be to eschew analysis of social media data altogether and instead focus on usage patterns that can be measured using traditional surveys. While I have yet to see any attempts to describe particular usage patterns as being more or less conducive to democratic governance, this is not to say we could not come up with such measures in the future.

One way might be to think about the relative proportion of people using social media platforms that are either directly controlled by the state (or else under the thumb of the state) as compared to people using private platforms. In the modern world, we might consider having access to social media platforms that are not subject to state censorship to be an important component of democratic accountability, and this would seem to be something that should be fairly easy to measure using survey data.[9]

A different type of approach might involve focusing on the segmentation of social media users into more or fewer different social media platforms. For example, if all conservatives use one social media network and all liberals another, this might be seen as worse (or better) for democratic survival than alternative arrangements.[10] A similar argument could be made regarding different ethnic groups in a country.

In both cases, the idea would be to use social media usage patterns to contribute to democracy scores. Validating such claims would of course involve a great deal of a priori work, but could provide another way in which social media could inform democracy measures.

6 Conclusions

As is often the case when discussing social media, there are both great challenges to and opportunities for utilizing social media in our measurement of democracy moving forward. Hopefully this essay has been useful in laying out a number of different paths by which such efforts could proceed, complete with both potential benefits and pitfalls.

I want to close with one final note: this essay has largely been predicated on the idea of developing cross-national measures of democracy. What is probably worth some serious consideration, however, is whether social media might be better suited to exploring within country cross-regional variation in the strength of democracy. This might be in an area in which we have fewer extant measures – thus offering more of a bang for the social media buck – but also might alleviate some of the concerns (e.g., working across different languages, massive data collection and retrieval efforts) in the cross-national context.

8. See Siegel et al. (2017), which examines the effect of being in more diverse social networks online in Egypt on the propensity to post tolerant – as opposed to intolerant – tweets.

9. Justifying such a claim is beyond the scope of this essay, but would likely involve similar arguments involving the value of a free press. As for the use of survey data, see for example the many excellent Pew Social Media Usage surveys in the United States, e.g. http://www.pewinternet.org/2016/11/11/social-media-update-2016/.

10. Consider, for example, Gab, a social media platform which Wired has called "the Alt-Right's Own Twitter" and "the ultimate filter bubble" (https://www.wired.com/2016/09/gab-alt-rights-twitter-ultimate-filter-bubble/)

Section News

Teorell, Jan. "Partisanship and Unreformed Bureaucracy: The Drivers of Election Fraud in Sweden, 1719-1908", Social Science History 41(2): 201-225, 2017.

Teorell, Jan & Michael Wahman. "Institutional stepping stones for democracy: how and why multipartyism enhances democratic change", Democratization, forthcoming (early view available here).

Gerardo Munck received the Frank Cass Prize for Best Overall Article inDemocratization, for "What is Democracy? A Reconceptualization of the Quality of Democracy," Democratization 23, 1 (2016).

Paul J. Carnegie recently moved from the Institute of Asian Studies, Universiti Brunei Darussalam to the University of the South Pacific where he is now Associate Professor in Political Science and Director of the Graduate Governance Program.

Lenka Bustikova and Cristina Corduneanu-Huci. 2017. "Patronage, Trust and State Capacity: The Historical Trajectories of Clientelism," World Politics, 69:2, 277-326.

Thomas E. Flores (Associate Professor, School for Conflict Analysis and Resolution, George Mason University) and Irfan Nooruddin (Hamad Bin Khalifa Professor of Indian Politics, School of Foreign Service, Georgetown University) published Elections in Hard Times: Building Stronger Democracies in the 21st Century (Cambridge University Press and Woodrow Wilson Center Press, 2016). The book argues that elections fail to promote democracy when held in countries with little democratic experience, scant fiscal space, and violent civil conflict.

Henry E. Hale and Timothy J. Colton, "Who Defects? Unpacking a Defection Cascade from Russia's Dominant Party 2008-12," American Political Science Review, v.111, no.2, May 2017, pp.322-37.

Henry E. Hale, "Russian Patronal Politics Beyond Putin," Daedalus, v.146, no.2, Spring 2017, pp.30-40, http://www.mitpressjournals.org/toc/daed/146/2

Henry Hale, "What It Means to Rule Like Putin," Seminar, issue 693, May 2017, pp.32-36, http://www.india-seminar.com/semframe.html

Call: Country Experts for V-Dem

Varieties of Democracy (V-Dem) is looking for Country Experts to assist with the next update in January 2018, willing to contribute to making the largest dataset on democracy even better in the future. In these times when democracy is at risk in many places, offer a few hours of your time, and contribute to the common good for science and international society!

V-Dem is a global, collaborative research project providing the largest ever database on democracy, headquartered at the V-Dem Institute, University of Gothenburg, Sweden. V-Dem provides distinct indices of electoral, liberal, participatory, deliberative, and egalitarian democracy, as well as some 40 indices of specific components of democracy and over 350 detailed indicators, for 177 countries over 117 years. The dataset contains 17 million data points, available for download or for online analysis free of charge on the V-Dem webpage: www.v-dem.net

The Role of a Country Expert
About half of the V-Dem data depends on evaluations by Country Experts - individuals with deep knowledge of a specific thematic area for a particular country, or set of countries.

We conduct annual updates of the V-Dem database. During these updates, most Country Experts work on one country, observed over the past several decades, and on one or two clusters of questions, e.g.: (1) elections and political parties, (2) the executive, the legislature and deliberation, (3) judiciary, civil liberty and sovereignty, and (4) civil society, media freedom and political equality. The coding procedure is entirely web-based. The default language for the online surveys is English but the survey also available in French, Spanish, Portuguese, Arabic and Russian.

Be a V-Dem Country Expert!

Section News

NEW BOOKS

A list of recently published books on democracy and democratization

Advanced Democracies

Democracy: A Case Study. By David A. Moss. Belknap Press of Harvard University Press, 2017. 773 pp.

When Ideas Mattered: A Nathan Glazer Reader. Edited by Joseph Dorman and Leslie Lenkowsky. Transaction Publishers, 2017. 352 pp.

Asia

Philippine Politics: Possibilities and Problems in a Localist Democracy. By Lynn T. White III. Routledge, 2015. 265 pp.

Purifying the Land of the Pure: A History of Pakistan's Religious Minorities. By Farahnaz Ispahani. Oxford University Press, 2017. 216 pp.

To Build a Free China: A Citizen's Journey. By Xu Zhiyong. Lynne Rienner, 2017. 297 pp.

When Crime Pays: Money and Muscle in Indian Politics. By Milan Vaishnav. Yale University Press, 2017. 410 pp.

Eastern Europe and the Former Soviet Union

Protest in Putin's Russia. By Mischa Gabowitsch. Polity, 2017. 332 pp.

Latin America and the Caribbean

Mano Dura: The Politics of Gang Control in El Salvador. By Sonja Wolf. University of Texas Press, 2017. 304 pp.

The Arab World Upended: Revolution and its Aftermath in Tunisia and Egypt. By David B. Ottaway. Lynne Rienner, 2017. 269 pp.

Egypt Beyond Tahrir Square. Edited by Bessma Momani and Eid Mohamed. Indiana University Press, 2016. 185 pp.

Comparative, Theoretical, General

Building Sustainable Peace: Timing and Sequencing of Post-Conflict Reconstruction and Peacebuilding. Edited by Arnim Langer and Graham K. Brown. Oxford University Press, 2016. 478 pp.

Corruption and Government Legitimacy: A Twenty-First Century Perspective. Edited by Johnathan Mendilow and Ilan Peleg. Lexington, 2016. 318 pp.

Honor in America? Tocqueville on American Enlightenment. By Laurie M. Johnson. Lexington, 2017. 155 pp.

Human Rights or Global Capitalism: The Limits of Privatization. By Manfred Nowak. University of Pennsylvania Press, 2017. 247 pp.

Peace Agreements: Finding Solutions in Intra-State Conflicts. By Nina Caspersen. Polity, 2017. 229 pp.

States in Disguise: Causes of State Support for Rebel Groups. By Belgin San- Akca. Oxford University Press, 2016. 320 pp.

The Rise and Fall of the Christian Myth: Restoring Our Democratic Ideals. By Burton L. Mack. Yale University Press, 2017. 310 pp.

APSA Annual Meeting 34-27 August, 2017: Panels sponsored by the Comparative Democratization Section

In August we meet again for the APSA Annual Meeting, this time held in the beautiful city of San Francisco.

The **Business Meeting** is held on **Friday, September 1 at 6:30p.m**. (Hilton, Franciscan C) and at our section **Reception at 7:30p.m**. (Hotel Nikko, Carmel II). Please join us to congratulate the winners of our section's five awards, to thank outgoing officers and to welcome incoming ones. Please let Melissa Aten (melissaa@ned.org) know if there are items that you would like to add to our agenda. We hope to see many of you there, and please note the terrific set of panels selected and organized by this year's section program chair Tom Flores. Thanks Tom!

Thursday, August 31, 2017

8:00 to 9:30am
Building the Leviathan: Institutions and Power in Authoritarian Regimes

8:00 to 9:30am
Who Supports (and Opposes) Democracy?

10:00 to 11:30am
To Coup or Not to Coup? The Military in Democratization

10:00 to 11:30am
Democratization in the Shadow of Civil War: Parties, Warriors, and Voters

12:00 to 1:30pm
Election Day Issues

2:00 to 3:30pm
Guardians of the Vote: Institutions for Building Electoral Integrity

2:00 to 3:30pm
Cat and Mouse: Smarter Authoritarians, Canny Oppositions

4:00 to 5:30pm
Democratic Abdication? Do Citizens Support the Subversion of Liberal Democracy?

Friday, September 1, 2017

8:00 to 9:30am
The Lines We Draw: Societal Divisions, Ethnicity, and Democratization

8:00 to 9:30am
Opposition Parties Under Authoritarianism

10:00 to 11:30am
Election Violence in Africa: Measurement, Causes, Consequences

12:00 to 1:30pm
Democratization by Elections?

12:00 to 1:30pm
Slip Slidin' Away: Why Do Democracies Backslide?

4:00 to 4:30pm
Poster Session: Comparative Democratization
In: Poster Sessions: New Political Science & Comparative Democratization

---

Saturday, September 2, 2017

8:00 to 9:30am
Loving the Big Man: Why Do Citizens Support Authoritarians?

8:00 to 9:30am
Promotion or Demotion? Understanding Democracy Promotion in Our World

8:00 to 9:30am
Here Today, Gone Tomorrow? Consolidating Young Democracies

10:00 to 11:30am
Global Populisms
Theme Panel
Roundtable

10:00 to 11:30am
Frustrated Hopes: Democratization in the Middle East

12:00 to 1:30pm
Taking it to the Streets: Protest, Civil Resistance, and Democratization

2:00 to 3:30pm
Making Democracy Work: Comparative Democratization in Brazil and South Africa

4:00 to 5:30pm
Peeling Back the Onion: Sub-National Variation in Democracy

4:00 to 5:30pm
Freedom and V-Dem: Lessons from the Varieties of Democracy Project(s)

APSA Annual Meeting 34-27 August

Sunday, September 3, 2017

8:00 to 9:30am
Subnational Democracy Within and Across Countries

10:00 to 11:30am
More than Words: Direct Citizen Participation in Democracy

Editorial Committee

APSA-CD is the official newsletter of the American Political Science Association's Comparative Democratization section. Formerly known as CompDem, it has been published three times a year since 2003. In October 2010, the newsletter was renamed APSA-CD and expanded to include substantive articles on democracy, as well as news and notes on the latest developments in the field. The newsletter is jointly produced and edited by faculty members of the V-Dem Institute.