KELLOGG INSTITUTE
FOR INTERNATIONAL STUDIES

working paper

A LEXICAL INDEX OF ELECTORAL DEMOCRACY

SVEND-ERIK SKAANING, JOHN GERRING
and HENRIKAS BARTUSEVIČIUS

399

June
2014

The Kellogg Institute for International Studies
University of Notre Dame
130 Hesburgh Center for International Studies
Notre Dame, IN 46556-5677
Phone: 574/631-6580
Web: kellogg.nd.edu

The Kellogg Institute for International Studies at the University of Notre Dame has built an international reputation by bringing the best of interdisciplinary scholarly inquiry to bear on democratization, human development, and other research themes relevant to contemporary societies around the world. Together, more than 100 faculty and visiting fellows as well as both graduate and undergraduate students make up the Kellogg community of scholars. Founded in 1982, the Institute promotes research, provides students with exceptional educational opportunities, and builds linkages across campus and around the world.

The Kellogg Working Paper Series:

- Shares work-in-progress in a timely way before final publication in scholarly books and journals
- Includes peer-reviewed papers by visiting and faculty fellows of the Institute
- Includes a Web database of texts and abstracts in English and Spanish or Portuguese
- Is indexed chronologically, by region and by research theme, and by author
- Most full manuscripts downloadable from kellogg.nd.edu

Contacts: Elizabeth Rankin, Editorial Manager
erankin3@nd.edu

A LEXICAL INDEX OF ELECTORAL DEMOCRACY

Working Paper #399 – June 2014

Svend-Erik Skaaning, John Gerring and Henrikas Bartusevičius

Svend-Erik Skaaning is professor of political science at Aarhus University where he does comparative politics research with a focus on the conceptualization, measurement, and explanation of democracy and human rights. In spring 2014 he was a visiting fellow with the Center for European Studies at Harvard University. He currently is working on two major research projects. The first, DEDERE, investigates the determinants of democratic stability in the interwar years. The second, Varieties of Democracy (V-Dem), examines the endogenous and exogenous explanations of political regime developments. Among his recent books are Requisites of Democracy (Routledge, 2011), Democracy and Democratization in Comparative Perspective (Routledge, 2013), and The Rule of Law (Palgrave Macmillan, 2014).

John Gerring is professor of political science at Boston University and a former visiting fellow at the Kellogg Institute for International Studies. He is the author of Party Ideologies in America, 1828–1996 (Cambridge University Press, 1998), Social Science Methodology: A Criterial Framework (Stanford University Press, 2001), Case Study Research: Principles and Practices (Cambridge University Press, 2007), A Centripetal Theory of Democratic Governance (Cambridge University Press, 2008), Concepts and Method: Giovanni Sartori and His Legacy (Routledge, 2009), Social Science Methodology: A Unified Framework, 2nd ed. (Cambridge University Press, 2012), An Applied Guide to Social Science Methodology (forthcoming, Cambridge University Press), along with numerous articles. He serves as co-PI of Varieties of Democracy (V-Dem) and the Global Leadership Project (GLP).

Henrikas Bartusevičius is a postdoctoral fellow with the DEDERE project at Aarhus University. In fall 2014 he starts a postdoctoral fellowship with the Conflict, Strategies, and Violence (CSV) project led by Kristian Skrede Gleditsch at the University of Essex. His research focuses on causes of civil conflicts, especially the effects of political institutions and socioeconomic inequalities. A visiting researcher at the Peace Research Institute Oslo (PRIO) in 2011, his work has appeared in the Journal of Peace Research.

ABSTRACT

Recent years have seen an efflorescence of work focused on the definition and operationalization of democracy. One debate concerns the scale, i.e., whether democracy is best measured by binary or graded scales. Critics of binary indices point out at that they are overly reductionist; all features of a regime must be reduced to a single coding decision, producing binary sets that lack discriminating power. Defenders counter that the different levels of graded measures are not associated with a specific set of conditions, meaning that they are difficult to interpret. Against this backdrop, we propose to operationalize electoral democracy as a series of necessary-and-sufficient conditions arrayed in an ordinal scale. The resulting "lexical" index of electoral democracy, based partly on new data collected by the authors, covers all independent countries of the world from 1800 to 2008. It incorporates binary coding of its sub-components based on factual characteristics of regimes and in this way reduces the problem of subjective judgments by coders for non-binary democracy indices. Binary codings are aggregated into an ordinal scale using a cumulative logic. In this fashion, we arrive at an index that performs a classificatory function – each level identifies a unique and theoretically meaningful regime type – as well as a discriminating function.

RESUMEN

En los últimos años hemos asistido a una erupción de trabajos enfocados en la definición y la operacionalización del concepto de democracia. Uno de los debates refiere a la escala, esto es, a si la mejor forma de medir la democracia es con escalas binarias o graduales. Los críticos de los índices binarios señalan que son excesivamente reduccionistas; todos los atributos de un régimen deben reducirse a una decisión de codificación singular, produciendo conjuntos binarios que carecen de poder de discriminación. Los defensores responden que los diferentes niveles de las medidas graduales no están asociados con conjuntos específicos de condiciones, lo que quiere decir que son difíciles de interpretar. Sobre este trasfondo, proponemos operacionalizar la noción de democracia electoral como una serie de condiciones tanto necesarias como suficientes organizadas en una escala ordinal. El índice "lexical" de democracia electoral resultante, basado en parte en nuevos datos recogidos por los autores, cubre todos los países independientes del mundo desde 1800 hasta 2008. Incorpora una codificación binaria de sub-componentes basada en características fácticas de los regímenes y de este modo reduce el problema de los juicios subjetivos de quienes codifican los índices de democracia no binarios. Las codificaciones binarias se agregan en una escala ordinal usando una lógica acumulativa. De este modo, llegamos a un índice que desempeña la función de clasificación –cada nivel identifica un tipo de régimen singular y teóricamente significativo—tanto como la función de discriminación.

Recent years have seen an efflorescence of work focused on the definition and operationalization of democracy. Because of its centrality to social science and its seeming recalcitrance, the concept has come to serve as a touchstone in methodological discussions of conceptualization and measurement (e.g., Goertz 2006; Schedler 2012; Seawright and Collier 2014). The field of extant indicators has been reviewed and critiqued many times (e.g., Armstrong 2011; Coppedge, Gerring, et al. 2011; Gleditsch and Ward 1997; Knutsen 2010; Munck 2009; Vermillion 2006), and a number of new entrants to the pantheon of global democracy indicators have been proposed.

One way to categorize this growing corpus of indicators is by the type of scale employed to measure the key concept, democracy—binary, ordinal, or interval. Binary indices include the Democracy-Dictatorship (DD) index produced by Przeworski and collaborators (Alvarez et al. 1996; Cheibub, Gandhi, and Vreeland 2010) and an index produced by Boix, Miller, and Rosato (2013, hereafter BMR). Ordinal measures include the Political Rights (PR) index and the Civil Liberty (CL) index, both produced by Freedom House (2013), along with the Polity2 index drawn from the Polity IV database (Marshall, Gurr, and Jaggers 2013). Interval measures include the Index of Democracy produced by Vanhanen (2000), the Contestation and Inclusiveness indices produced by Coppedge, Alvarez, and Maldonado (2008), and the Unified Democracy Scores (UDS) produced by Pemstein, Meserve, and Melton (2010).

Evidently, there is much more to a democracy index than its choice of scale. Yet the latter is a critical issue in measurement and one that has garnered considerable controversy, especially as concerns the virtues and vices of binary measures (contrast Elkins 2000 and Cheibub, Gandhi, and Vreeland 2010). Critics of binary indices point out their reductionist elements: all features of a regime must be reduced to a single coding decision, producing binary sets that are highly heterogeneous and borderline cases that may not fit neatly into either category. Binary indices, by construction, lack discriminating power. Defenders counter that if the definition of these binary sets is properly grounded in theory, the two-part typology may succeed in identifying—from the multifarious elements of democracy—that condition, or set of conditions, that serves a crucial role in political life (see Collier and Adcock 1999). However, this is not an easy claim to

sustain—witness the proliferation of binary indices that identify different defining conditions of democracy.[1]

We take for granted that different sorts of scales are useful for different purposes. Our aim is thus not to subsume or replace extant measures of democracy. The discipline is well served by a variety of measures of this central concept. Rather, we propose a new method of scale construction that overcomes some of the perceived deficits of extant scales.

Specifically, we propose to operationalize electoral democracy as a series of necessary-and-sufficient conditions arrayed in an ordinal scale. We refer to this scaling procedure as "lexical." The resulting Lexical index of electoral democracy, partly based on novel data construction, covers all independent countries of the world in the period 1800–2008 and is thus the most comprehensive measure of democracy available in terms of coverage. It incorporates binary coding based on factual characteristics of regimes and in this way avoids the problem of subjective judgments by coders and the "mashup" quality (Ravallion 2011) of non-binary democracy indices. However, each binary coding is aggregated together using the cumulative logic of a lexical scale with seven levels. In this fashion, we arrive at an index that performs a classificatory function—each level identifies a unique regime type—as well as a discriminating function. This approach to measurement offers theoretical and empirical advantages over other methods of representing the complex concept of electoral democracy that may be useful in certain settings.

The first section of the paper shows that extant data sets of democracy fall short in simultaneously providing fine-grained discriminatory power and meaningful categories. The second section develops conditions that define our Lexical index. The third section discusses how this index is coded across the universe of independent states from 1800 to 2008. The fourth section explores features of the Lexical index, which is compared with extant indices in the fifth section. The sixth section applies the new measure to the question of state repression, showing how its fixed meanings to the different levels inform the interpretation of statistical relationships in a way that is not accessible through conventional democracy indices. We conclude with a few observations about the implications and purview of this new index of electoral democracy.

[1] A short list would include DD and BMR—already discussed—as well as Bernhard, Nordstrom, and Reenock (2001) and Persson and Tabellini (2006).

DISCRIMINATION VS. MEANINGFUL CATEGORIES

The Freedom House indices recognize seven categories (fourteen if they are combined into a single index), and the Polity2 index twenty-one categories.[2] In contrast to binary indices, the levels in these ordinal indices are not qualitatively different from each other. A "3" on the PR, CL, or Polity2 scale signifies that a polity is more democratic than a country coded as "2," but it does not signify clearly identifiable traits that distinguish the two polities. Ordinal scales of democracy identify countries with more or less democracy but not different kinds of democracy. This has probably contributed to the tendency to employ these indices as if they possessed the measurement attributes of an interval scale.

Interval indices of democracy are generally second-order indices. That is, they are constructed by aggregating together information provided by other democratic indices—through factor analysis (Contestation and Inclusiveness) or IRT models (UDS). The exception is Vanhanen's Index of Democracy. However, the distribution of data on this index is so highly skewed and so evidently censored—nearly 50 percent of the observations are at the zero point of a 100-point scale—that its claims to be an interval scale measure are compromised. Thus, our discussion of interval measures will focus on the Contestation, Inclusiveness, and UDS indices.

The purpose of a well-constructed interval index is to identify fine distinctions among entities. The Contestation, Inclusiveness, and UDS indices achieve this goal as well as can be expected. However, the goal of reducing the plenitude of characteristics associated with "democracy" to a single unidimensional index is elusive. It is elusive because the concept itself is multidimensional and because extant indicators are limited in their purview (Coppedge, Gerring, et al. 2011). An appropriate response is to define the resulting index in a carefully delineated way—as representing only one dimension of a multifaceted concept. Thus, Coppedge, Alvarez, and Maldonado (2008) describe one component from their principal components analysis as Contestation and the other as Inclusiveness. The UDS is simply described as a measure of democracy. However, these are ex post descriptions resulting from a rather ad hoc process. They result from putting together myriad indices—whose definition and construction is itself something of a mystery—with a statistical model and labeling the central tendencies resulting from that model as "X." It is unclear, in other words, whether the index labeled

[2] The Polity2 index is less discriminating than it appears; countries tend to bunch in two areas—toward the bottom of the index and at the very top—producing a strongly bimodal distribution

"Inclusiveness" includes all measures relevant to that concept and no measures irrelevant to that concept and whether the included elements are aggregated in an appropriate way. Evidently, principal components analysis is only one of many possible approaches to the task of aggregation. Although the menu of choices is limited to a small class of measurement models—pca, IRT, and the like—in fact, the array of possibilities is virtually limitless, given that researchers must make many choices in the construction of a given IRT model.

If all indices are in some sense arbitrary, they are arbitrary in strikingly different ways. The arbitrariness of a binary scale lies in the choice of necessary condition(s) that define the two categories. The arbitrariness of an ordinal or interval scale lies in the choice of indicators to include as elements of the index and the choice of aggregation method to combine those indicators into a single scale.

Likewise, if all indices are informative, they are informative in strikingly different ways. The information contained in a binary index is classificatory, that is, how to group polities in a fashion that is theoretically and empirically fecund. The information contained in an interval index is discriminatory, that is, how to identify small differences between polities that are similar to each other. Ordinal scales occupy a middle position in this respect. However, extant ordinal indices of democracy perform neither task very well, for the reasons explained above.

DEVELOPING A LEXICAL INDEX

Our index of democracy focuses explicitly on the electoral aspect of democracy, the idea that democracy is achieved through competition among leadership groups that vie for the electorate's approval during periodic elections before a broad electorate. This view of our subject may also be referred to as the competitive, elite, minimalist, procedural, realist, 'thin', or Schumpeterian conception of democracy (Przeworski et al. 2000; Schumpeter 1950; Møller and Skaaning 2013).

Broadly interpreted, this understanding of democracy is consistent with the construction of most extant indices of democracy. However, it is worth stressing that our definition of electoral democracy is narrower than most others—especially extant ordinal and interval indices (PR, CL, Polity2, Contestation, Inclusiveness, UDS). We are concerned with the degree to which political authority is subject to periodic elections. We are not concerned with other aspects of democracy such as civil liberties, rule of law, constraints on executive power, deliberation, or

non-electoral mechanisms of participation—except insofar as any of these might affect the quality of elections. Electoral refers to elections.

A lexical approach to measurement is concept-driven (Gerring, Skaaning, and Pemstein 2014). Thus, we begin with a survey of attributes associated with the key concept, electoral democracy. In identifying attributes for possible inclusion in our index we are mindful of the vast literature on this topic, with special attention to linguistic studies of the concept (e.g., Held 2006; Lively 1975; Naess, Christophersen, and Kvalo 1956) and foundational works in the electoral tradition (as listed above).

To form a lexical scale one must arrange attributes so that each serves as a necessary-and-sufficient condition within an ordered scale. That is, each successive level is comprised of an additional condition, which defines the scale in a cumulative fashion. Condition A is necessary and sufficient for L1; conditions A&B are necessary and sufficient for L2; and so forth. In achieving these desiderata four criteria must be satisfied: (1) binary values, (2) unidimensionality, (3) qualitative differences, and (4) centrality or dependence (see Gerring, Skaaning, and Pemstein 2014).

First, each level in the scale must be measurable in a binary fashion without recourse to arbitrary distinctions. It is either satisfied or it is not. To be sure, the construction of a binary condition may be the product of a set of necessary and/or sufficient conditions. Collectively, however, these conditions must be regarded as necessary and sufficient.

Second, levels in a lexical scale must be understood as elements of a single latent (unobserved) concept. Conceptual multidimensionality must be eliminated, either by dropping the offending attribute and/or by refining the concept in a clearer and perhaps more restrictive fashion, as we have in moving from "democracy" to "electoral democracy."

Third, each level must demarcate a distinct step or threshold in a concept, not simply a matter of degrees. Levels in a lexical scale are intended to identify qualitative differences. A "3" on a lexical scale is not simply a midway station between "2" and "4." Indeed, each level may be viewed as a subtype of the larger concept. (Note that these subtypes are defined by cumulative combinations of the attributes possessed by the full concept—A, A&B, A&B&C, and so forth—fulfilling the criterion of a classical concept.)

The most challenging aspect of lexical scale construction is the ordering of attributes, which follows a conceptual (rather than empirical) logic. One attribute may be considered prior

to another if it is more central to the concept of theoretical interest (from some theoretical vantage point). This follows a constitutive approach to measurement, where attributes are the defining elements of a concept. Alternatively, one attribute may be considered prior if it is a logical, functional, or causal prerequisite of another. The dependence of B on A is what mandates that A assume a lower level on a scale. Whether responding to considerations of centrality or dependence, the levels of a lexical scale bear an asymmetric relationship to each other; some are more fundamental than others. This is the most distinctive feature of a lexical scale.[3]

Accordingly, we attempt to gauge the relative centrality or interdependence of various attributes of electoral democracy. The existence of elections is judged fundamental (conditio sine qua non), since no other attributes commonly associated with electoral democracy make any sense outside of an electoral context (Collier and Adcock 1999: 559; Merkel 2004: 36–38). One could not say, for example, that Country A is more of an electoral democracy than Country B if neither polity holds elections, regardless of what other characteristics those polities might possess. Likewise, some attributes depend upon other attributes in a logical manner. For example, one cannot hold multiparty elections unless an electoral regime is in place; elections are a necessary condition of multiparty elections. Finally, some of the attributes of the concept depend for their meaning on other attributes in a functional manner. For example, the extent of suffrage allowed in an election is of little import unless elections count for something (i.e., unless they allow for multiparty competition and the most important policymaking offices are elective). Whether or not suffrage is restricted or universal in North Korea is not a question that is generally regarded as consequential. (North Korea would hardly be less democratic if it decided to restrict access to the ballot, while changing no other aspect of its totalitarian system.)

Based on these considerations, we arrive at a Lexical index of electoral democracy with six conditions and seven levels, as follows:

[3] Where lexical ordering is unclear *a priori* (according to considerations of centrality and dependence), one is well advised to consider the shape of the empirical universe. Specifically, if *A* is always (or almost always) present where *B* is present, there may be grounds for considering *A* as more central or more fundamental than *B*. However, any conclusions reached on the basis of an exploration of empirical properties must be justified as a matter of centrality or dependence. Thus, we regard the relative prevalence of attributes as a clue to asymmetric relationships among the properties of a concept, not as a desideratum. In constructing a lexical scale, deductive considerations trump data distributions.

0. No elections.

1. No-party or one-party elections.

2. Multiparty elections for legislature.

3. Multiparty elections for legislature and executive.

4. Minimally competitive, multiparty elections for legislature and executive.

5. Minimally competitive, multiparty elections with full male or female suffrage for legislature and executive.

6. Minimally competitive, multiparty elections with universal suffrage for legislature and executive.

Further elaboration of this minimalist approach to electoral democracy can easily be envisioned. One might, for example, include an eighth level measuring aspects of electoral integrity (i.e., no irregularities with respect to intimidation, disruption, violence, and registration, and high respect for political liberties such as freedom of expression, assembly, and association). For present purposes, we restrict ourselves to what might be considered the most basic properties of electoral democracy.

Recall again that electoral democracy identifies one aspect, or dimension, of the parent concept. Other dimensions of democracy—sometimes articulated as liberal, deliberative, egalitarian, or participatory—are ignored, though presumably lexical indices might also be developed for these concepts.

CODING

To operationalize the levels of the index we make use of four variables from the Political Institutions and Events (PIPE) dataset (Przeworski et al. 2013), defined as follows:

> LEGSELEC: If there is a legislative body that issues at least some laws and does not perform executive functions and (the lower house of) the legislature is at least partly elected and has not been closed, the variable is scored 1; otherwise 0.
>
> EXSELEC: If the chief executive is either directly elected or indirectly elected (i.e., chosen by people who have been elected), the variable is scored 1; otherwise 0.

OPPOSITION: If there is an legislature that is at least in part elected by voters facing more than one choice, meaning that all candidates at elections are not presented on the same, single list, there is not only one party while some candidates run as independents, parties are not generally banned and everyone run without party labels, and the legislature has not been dissolved, this variable is scored 1; otherwise 0.

FRANCHISE: This variable is divided into two components, which we refer to as MALE SUFFRAGE and FEMALE SUFFRAGE. If there is virtually universal suffrage for male and female citizens, respectively, for national elections, these variables are scored 1; otherwise 0. Restrictions referring to age, criminal convictions, legal incompetence, and local residency are not considered as violations of the suffrage criterion.

Since the quality of elections is not taken into account in the foregoing variables, we generate a new variable, as follows:

COMPETITION: The chief executive offices and the seats in the effective legislative body are filled by elections characterized by uncertainty, meaning that the elections are, in principle, sufficiently free to enable the opposition to gain power. If control over the key executive and legislative offices is determined, in practice, by the electorate by means of contested elections, the executive and members of the legislature have not been unconstitutionally removed, the legislature has not been dissolved, (non-extremist) parties have not been banned, and the constitutional timing of elections is not violated in a more than marginal fashion, the variable is scored 1; otherwise 0.

This indicator captures whether or not elections are contested, that is, whether there is a positive probability that the opposition can win government power (see Przeworski et al. 2000: 15–18; Møller and Skaaning 2013). In common with the DD and the BMR indices, we consider instances of incumbent turnover (as a result of elections) as a strong indicator of contested elections. However, we do not consider executive turnover to be either necessary or sufficient for genuinely

contested elections.[4] It should be clear that in classifying an election as competitive we are establishing a modest threshold, not insisting on an entirely level playing field or a high level of respect for civil liberties. Specifically, elections are considered competitive if the winner of executive and legislative elections reflects the votes cast by the electorate, as near as can be determined (from extant sources). This state of affairs is reached if opposition candidates are generally free to run for these key offices, voters experience little systematic coercion in exercising their electoral choice, and electoral fraud does not determine who wins.

A country-year is assigned the highest score (0–6) for which it fulfills all the requisite criteria, as follows:

0. LEGSELEC=0 & EXSELEC=0.

1. LEGSELEC=1 or EXSELEC=1.

2. LEGSELEC=1 & OPPOSITION=1.

3. LEGSELEC=1 & OPPOSITION=1 & EXSELEC=1.

4. LEGSELEC=1 & OPPOSITION=1 & EXSELEC=1 & COMPETITION=1.

5. LEGSELEC=1 & OPPOSITION=1 & EXSELEC=1 & COMPETITION=1 & (MALE SUFFRAGE=1 or FEMALE SUFFRAGE=1).[5]

6. LEGSELEC=1 & OPPOSITION=1 & EXSELEC=1 & COMPETITION=1 & MALE SUFFRAGE=1 & FEMALE SUFFRAGE=1.

[4] It is not necessary since an incumbent party can be sufficiently popular to win a long sequence of genuinely contested elections, as happened for decades in, e.g., Botswana, Japan, and Sweden. It is not sufficient because opposition power can gain power through a flawed election if the incumbents have only weak control on power or have stepped down. That said, in all but a very few cases executive turnover in conjunction with elections is associated with a coding of 1 for COMPETITION.

[5] In no extant cases was universal female suffrage introduced before universal male suffrage, so in practice this level is reserved for countries with male (only) suffrage.

Countries are coded for the length of their sovereign existence within the 1800–2008 timespan, generating a dataset with 220 countries and 17,169 country-years. To identify independent countries we rely on Gleditsch (2013) and Correlates of War (2011), supplemented from 1800 to 1815 by various country-specific sources. Importantly, electoral democracy does not presume complete sovereignty. A polity may be constrained in its actions by other states, by imperial control (as over a colony), by international treaties, or by world markets. Thus, to say that a polity is an electoral democracy is to say that it functions as such for policies over which it enjoys decision-making power.

Scores for each indicator reflect the status of a country on the last day of the calendar year (31 December) and are not intended to reflect the mean value of an indicator across the previous 364 days.

To qualify as an election the electorate may be quite small—though it must be separable from, and much larger than, the group of officials it is charged with selecting. Examples would be South Africa under Apartheid and virtually all national elections in Europe during the nineteenth century. In measuring universal male and female suffrage we take a juridical approach. Suffrage is achieved when constitutionally prescribed, even though local or informal practices may impede the achievement of this right (as in the American South prior to the Civil Rights movement). This is consistent with the usage of the concept by Schumpeter and Przeworski and also with many extant indices (e.g., Polity2 and DD). Indirect elections do not qualify as "elections" unless the electors endorse specific candidates or parties, as in US presidential elections.

Although we employ PIPE as an initial source for coding LEGSELEC, EXSELEC, OPPOSITION, and FRANCHISE (MALE SUFFRAGE and FEMALE SUFFRAGE), we deviate from PIPE codings—based on our reading of country-specific sources—in several ways. First, with respect to executive elections, in the PIPE dataset "Prime ministers are always coded as elected if the legislature is open." However, for our purposes we need an indicator that also takes into account whether the government is responsible to an elected parliament if the executive is not directly elected—a situation generated by a number of European monarchies prior to World War I, by episodes of international supervision such as Bosnia-Herzegovina in the first years following the civil war, and by some monarchies in the Middle East and elsewhere (e.g., Liechtenstein, Monaco, and Tonga) in the contemporary era. To illustrate, PIPE codes Denmark

as having executive elections from 1849 to 1900 although the parliamentary principle was not established until 1901. Before then, the government was accountable to the king. Among the current cases with elected multiparty legislatures not fulfilling this condition, we find Jordan and Morocco. In order to achieve a higher level of concept-measure consistency, we have thus recoded all country-years (based on country-specific accounts) for this variable where our sources suggested doing so.

Moreover, we complete all missing values (and missing countries, e.g., the German principalities of the nineteenth century) in the PIPE dataset, generating a complete dataset for all conditions for all independent countries of the world in the period under study (1800–2008). Whereas the numbers of observations for the employed PIPE indicators range between 14,465 and 15,302, our dataset provides 17,179 observations for all indicators. Except for minor adjustments regarding executive elections (mentioned above), this additional coding follows the rules laid out in the PIPE codebook. Coding decisions are based on country-specific sources that are too numerous to specify. In rare instances we stumbled upon information that required a recoding of PIPE variables, so the two datasets do not correspond exactly.

When contrasted with most continuous measures of democracy, the Lexical index is relatively simple, enhancing its transparency and reproducibility. Coding decisions are factual in nature, resting on institutional features that require historical knowledge but not subjective judgments on the part of the coder. To be sure, uncertainties are introduced when source material for a country is weak. But we assume that this sort of bias is random rather than systematic (as it might be if coder judgments involved questions of meaning and interpretation). In this respect, the Lexical index echoes a feature of most binary indices (e.g., DD and BMR). Indeed, it is quite similar to these indices insofar as it relies on binary codings, which are combined to form a cumulative index.

Another important feature of the coding procedure is its separability from other factors that sometimes confound our ability to measure political institutions. When coding democracy (D) and governance (G) indices—particularly those that assume a continuous distribution—there is a strong possibility that coders may view the state of D or G in Country X as inseparable from the general state of affairs in that country, including its economic performance. When things are going well, X may receive a higher score. When things are going poorly, it may receive a lower score, even if its political institutions are substantially unchanged (Kurtz and Schrank 2007). The

coding of the Lexical index offers little opportunity for this sort of measurement error because coding decisions rest on clear-cut thresholds and because the features that are being coded are not amenable to "state of affairs" confounders.

To provide an empirical check on reproducibility we conducted an inter-coder reliability test. By design, one of the authors (Bartusevičius) was not involved in the construction of the index or the original coding of the dataset and was not informed of codings arrived at by the other authors or by the PIPE dataset. He was then assigned the task of re-coding twenty-two countries (10 percent of the sample), chosen at random, based on the coding rules presented above and using only country-specific sources (which he chose based upon his review of the extant literature).

TABLE 1

INTER-CODER RELIABILITY

	Agreement (%)	Cohen's Kappa	Krippendorff's Alpha
LEGSELEC	93.72	.831	.831
EXSELEC	95.63	.903	.903
OPPOSITION	94.54	.888	.888
MALE SUFFRAGE	95.49	.898	.898
FEMALE SUFFRAGE	96.38	.926	.926
COMPETITION	96.66	.920	.920
LEXICAL INDEX	87.37	.840	.943

Three standard statistical measures of inter-coder reliability are presented in Table 1: percent agreement, Cohen's kappa, and Krippendorff's alpha. These are calculated at the variable level (for LEGSELEC, EXSELEC, OPPOSITION, MALE SUFFRAGE, FEMALE SUFFRAGE, and COMPETITION) and at the composite level (for the Lexical index). All measures report high levels of inter-coder reliability, suggesting that the index is readily reproducible.

FEATURES OF THE DATASET

In order to get a feel for the application of the Lexical index, we provide country scores for the median year in our sample, 1904, as shown in Table 2. At that time, there were fifty-three independent countries in the world. These were distributed fairly evenly across the seven categories of the Lexical index, with the exception of the most democratic category (6), which

has only one occupant. Only Australia granted universal suffrage to both men and women, while satisfying the other criteria stipulated in the index. (New Zealand—often considered as the first country to introduce universal suffrage—did not become independent before 1907 according to our criteria.)

TABLE 2

FREQUENCY DISTRIBUTION, 1904

0 - **National elections**	1 + **National elections**	2 + **Multiparty elections for legislature**	3 + **Multiparty elections for executive**	4 + **Competitive elections**	5 + **Male or female suffrage**	6 + **Universal suffrage**
Afghanistan	*Colombia*	*Austria-Hungary*	*Argentina*	*Chile*	*Belgium*	*Australia*
China	*Cuba*	*Bulgaria*	*Bolivia*	*Costa Rica*	*Canada*	
Ethiopia	*Haiti*	*Germany*	*Brazil*	*Denmark*	*France*	
Iran	*Honduras*	*Japan*	*Dominican Rep.*	*Luxembourg*	*Greece*	
Korea	*Liberia*	*Portugal*	*Ecuador*	*Netherlands*	*Switzerland*	
Montenegro	*Mexico*	*Romania*	*El Salvador*	*Panama*	*United States*	
Nepal	*Nicaragua*	*Spain*	*Guatemala*	*United Kingdom*		
Oman	*Paraguay*	*Sweden*	*Italy*	*Uruguay*		
Ottoman Emp.	*Venezuela*		*Peru*			
Russia			*Serbia*			
Thailand						

N=53

A frequency distribution of scores across the entire 1800–2008 period is provided in Table 3. It will be seen that the most populated categories are L0, L1, L3, and L6, while others (notably L5) have fewer occupants. A fairly high proportion of cases stack up at the two ends of the index, a problem that affects many ordinal and interval indices (Cheibub, Gandhi, and Vreeland 2010: 77; Treier and Jackman 2008).

TABLE 3

FREQUENCY DISTRIBUTION, 1800–2008

	N	%
0	4,500	26.21
1	2,909	16.94
2	1,484	8.64
3	2,618	15.25
4	864	5.03
5	503	2.93
6	4,291	24.99

The distribution of cases changes over time, as one might expect. This feature may be portrayed in a stacked graph of the categories in each year, as shown in Figure 1. Note that our sample grows over time—from 27 in 1800 to 194 in 2008—due to the appearance of newly sovereign states (e.g., in Africa) and the break-up of sovereign states (e.g., the Soviet Union).

FIGURE 1

RELATIVE AND ABSOLUTE DISTRIBUTION OF COUNTRIES, 1800–2008

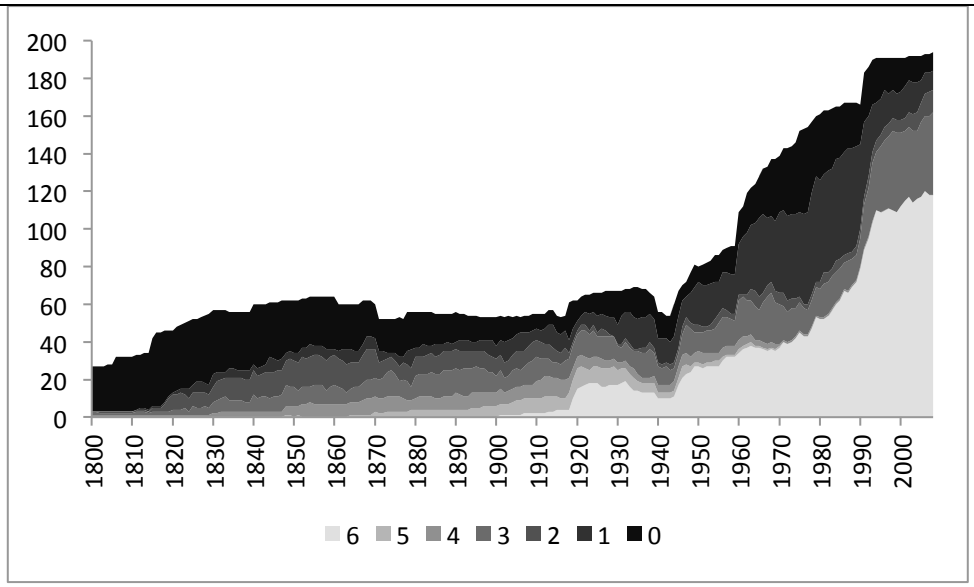

Figure 1 also illustrates a key feature of a lexical scale, namely the possibility of decomposing a concept into constituent parts—parts that are intrinsically meaningful because they represent qualitative (i.e., step or threshold) differences. Each lexical scale contains an implicit typology. This allows one to interpret category membership—and changes in membership over time—in a meaningful way. For example, one can evaluate the composition of political regimes at each point in time over the past two centuries.

In 1800 polities were predominantly of type 0 (no elections). Later in the nineteenth century we see the rise of types 1–5 and the concomitant decline of type 0. This is the most diverse period, when no single type is dominant, as illustrated by our snapshot of the world in 1904 (see Table 2). Over the course of the twentieth century, we can see the extraordinary rise of type 1 (elections without multiparty competition), followed by a steep decline beginning in the 1980s and coincident with the Third Wave of democratization (Huntington 1991).

Type 3 (multiparty executive and legislative elections without real competition), which had been a modestly sized category for a century, begins to grow in the late twentieth century to the point where it constitutes the second-most dominant regime-type. This regime-type is quite similar to polities described as competitive authoritarian (Levitsky and Way 2002) or limited multiparty (Hadenius and Teorell 2007).

The other striking pattern over the past century is the rise of type 6, the highest level of our Lexical index, corresponding to polities that satisfy all assessed criteria. This category now comprises over half of all polities in the world.

Figure 1 also offers insight into those periods in which electoral democracy advanced throughout the world (e.g., at the end of World War I, World War II, and the Cold War), as well as those periods in which it declined (e.g., the 1930s).

CONTRASTS WITH EXTANT INDICES

Table 4 summarizes salient features of the Lexical index alongside the nine extant democracy indices introduced at the outset. It will be seen that the Lexical index has much broader historical coverage than DD, PR, CL, Contestation, Inclusiveness, and UDS—all of which are focused on the contemporary era—and slightly better coverage than Polity2, BMR, and Vanhanen's Democracy index.

TABLE 4

INDICES COMPARED

INDEX	SCALE		COVERAGE			CORRELATION (with Lexical)	
	Type	Range	Countries	Years	Obs	All obs	Lexical<6
Lexical (authors)	Lexical	0–6	220	1800–2008	17,169		
DD (Cheibub et al.)	Binary	0–1	197	1946–2008	9,115	.83 (S)	.37 (S)
BMR (Boix et al.)	Binary	0–1	210	1800–2007	16,308	.79 (S)	.45 (S)
Polity2 (Marshall et al.)	Ordinal	-10–10	185	1800–2012	16,327	.79 (S)	.59 (S)
PR (Freedom House)	Ordinal	1–7	199	1972–2012	7,040	.85 (S)	.43 (S)
CL (Freedom House)	Ordinal	1–7	199	1972–2012	7,040	.79 (S)	.31 (S)
Democracy Index (Vanhanen)	Interval	0–100*	192	1810–2010	15,149	.78 (P)	.49 (P)
Contestation (Coppedge et al. 2008)	Interval	-1.84–1.96	197	1950–2000	7,534	.90 (P)	.66 (P)
Inclusiveness (Coppedge et al. 2008)	Interval	-3.04–1.91	197	1950–2000	7,534	.59 (P)	.54 (P)
UDS (Pemstein et al.)	Interval	-2.11–2.26	199	1946–2012	9,850	.87 (P)	.63 (P)

*=theoretical range. S = Spearman's rho. P = Pearson's correlation coefficient. Country counts are based on COW country codes (extended with additional, unique country codes for Orange Free State, Transvaal, Tibet, and United Provinces of Central America, as suggested by Gleditsch), whereas years and observations are taken from the original datasets.

As is to be expected, the Lexical index generally co-varies with other indices. For example, it correlates with Polity2 at 0.79 and with the Political Rights index at 0.85 (Spearman's rho). However, when the highest scoring cases (Lexical=6) are dropped from the sample these correlations drop to 0.59 and 0.43, respectively.

A more detailed look at the relationships between extant binary (DD, BMR) and ordinal (PR, Polity2) indices and the Lexical index is portrayed in cross-tabulations in Table 5. This confirms that while various measures of electoral democracy are related, they are not very highly correlated.

TABLE 5

CROSS-TABULATIONS

Lexical

		0	1	2	3	4	5	6	Sum
DD	**0**	1,351	1,969	294	1,184	50	2	259	5,109
	1	0	13	27	187	101	48	3,567	3,943
	Sum	1,351	1,982	321	1,371	151	50	3,826	9,052
BMR	**0**	3,788	2,880	1,245	2,444	418	63	164	11,002
	1	7	10	42	78	495	390	3,907	4,929
	Sum	3,795	2,890	1,287	2,522	913	453	4,071	15,931
PR	**0**	397	478	17	87	0	0	0	979
	1	330	488	39	279	1	0	0	1,137
	2	144	200	67	238	15	0	24	688
	3	20	54	49	224	4	0	173	524
	4	3	16	12	66	0	0	395	492
	5	2	0	17	26	0	0	860	905
	6	0	0	17	1	0	0	1,501	1,519
	Sum	896	1,236	218	921	20	0	2,953	6,244
Polity2	**0**	1,118	119	39	0	0	0	0	1,276
	1	252	609	114	133	0	0	4	1,112
	2	126	212	76	76	0	1	0	491
	3	640	765	205	105	0	0	0	1,715
	4	676	222	158	197	0	0	1	1,254
	5	145	161	40	197	20	0	9	572
	6	51	79	258	183	24	2	1	598
	7	241	267	85	414	45	42	15	1,109
	8	16	23	34	147	39	1	3	263
	9	56	132	52	243	6	4	22	515
	10	109	52	10	129	8	9	19	336
	11	145	31	76	95	8	7	3	365
	12	19	39	44	191	66	21	21	401
	13	6	25	9	77	76	14	37	244
	14	87	10	24	121	161	8	78	489
	15	3	26	11	90	50	19	164	363
	16	6	3	27	19	53	22	283	413
	17	4	13	0	36	67	37	300	457
	18	6	12	4	24	51	31	500	628
	19	0	1	1	4	50	56	377	489
	20	3	2	0	10	187	176	1,756	2,134
	Sum	3,709	2,803	1,267	2,491	911	450	3,593	15,224

PR and Polity2 rescaled so that 0 = lowest value.

One might infer that the Lexical index is an outlier among democracy indices. However, a principal components analysis, shown in Table 6, reveals that this is not the case. Again, we find a striking contrast between full sample and partial sample results. In the full sample, 83 percent of the variance across these ten indices is explained by the first component. In the partial sample (Lexical<6), only 52 percent of the variance can be explained by the first component. However, in neither analysis is the Lexical index an outlier, as shown in the eigenvalues.

TABLE 6

PRINCIPAL COMPONENTS ANALYSIS

	Full sample				Partial sample			
Component	**Eigenvalue**	**Difference**	**Proportion**	**Cumulative**	**Eigenvalue**	**Difference**	**Proportion**	**Cumulative**
1	8.3170	7.5898	0.8317	0.8317	5.1723	3.8357	0.5172	0.5172
2	0.7272	0.3974	0.0727	0.9044	1.3366	0.1209	0.1337	0.6509
3	0.3298	0.1112	0.0330	0.9374	1.2158	0.5377	0.1216	0.7725
4	0.2186	0.0860	0.0219	0.9593	0.6781	0.2448	0.0678	0.8403
5	0.1326	0.0364	0.0133	0.9725	0.4333	0.0307	0.0433	0.8836
6	0.0963	0.0272	0.0096	0.9821	0.4026	0.1133	0.0403	0.9239
7	0.0691	0.0088	0.0069	0.9890	0.2893	0.0367	0.0289	0.9528
8	0.0603	0.0312	0.0060	0.9951	0.2526	0.1089	0.0253	0.9781
9	0.0292	0.0091	0.0029	0.9980	0.1437	0.0679	0.0144	0.9924
10	0.0201	.	0.0020	1.0000	0.0758	.	0.0076	1.0000

EIGENVECTORS

	Full sample		Partial sample	
Variable	**Comp1**	**Unexplained**	**Comp1**	**Unexplained**
Lexical	0.3282	0.1040	0.3222	0.4631
DD	0.3143	0.1786	0.2033	0.7861
BMR	0.3215	0.1406	0.1533	0.8785
Polity2	0.3314	0.0865	0.3543	0.3508
PR	0.3322	0.0820	0.3589	0.3337
CL	0.3168	0.1654	0.3165	0.4820
Democracy index	0.3153	0.1731	0.3308	0.4339
Contestation	0.3406	0.0349	0.4074	0.1413
Inclusiveness	0.2016	0.6618	0.1828	0.8272
UDS	0.3369	0.0562	0.4099	0.1311

In elucidating the distinctive features of our Lexical index a useful point of comparison is provided by binary indices. The latter generally combine several of the features identified in our ordinal scale. For example, DD may be said to combine L1–4 while BMR combines L1–5, with suffrage understood as a majority of men rather than all men. In doing so, the authors suggest that a polity cannot be called an electoral democracy until it has satisfied a number of conditions—though these conditions do not exactly map onto the condition utilized to score the Lexical index, as shown in Table 5.

Our index does not take issue with this determination. However, a lexical approach to scaling suggests that polities that fail to pass all four or five of these conditions may nonetheless be regarded as partial members of the class "electoral democracy." For example, it suggests that a polity with elections is closer to the electoral ideal than a polity without elections. And it suggests that this distinction—along with others identified along the seven-level index—has consequences, consequences that can be understood as greater/lesser possession of various traits associated with electoral democracy. Thus, rather than insisting that a number of necessary conditions be met, we regard each condition as providing a threshold on a single ordinal scale.

Clearly, the Lexical index allows one to represent more information than is possible in a binary scale. At the same time, the sensitivity of a seven-level ordinal scale is less than a longer ordinal scale (e.g., Polity) and much less than an interval scale (e.g., UDS). In terms of discriminatory ability, the Lexical index occupies a midway point.

The advantage of lexical scaling relative to more differentiated ordinal scales or interval scales is in clarity. While the latter are derived from complex models (e.g., UDS) or less formulaic but often opaque weightings across dimensions (e.g., Freedom House and Polity), the lexical value affixed to a country in a particular year is immediately interpretable. We know what a "5" means because there is only one combination of attributes that will yield a score of 5 on a lexical scale.

Likewise, we can understand the categories of the scale as indicating discrete polity types, which can be tracked through time, as in Figure 1. By way of contrast, longer ordinal scales (e.g., Polity) and interval indices (e.g., UDS) allow one to track the overall trends—more or less democracy through time—by examining changes in the mean over time. But they cannot indicate anything about the specific content (quality) of regimes or about which regime-types

expanded or contracted at different points in time. The latter information is both substantively important as well as useful for tracing causal mechanisms, as discussed below.

THE LEXICAL INDEX AT WORK: DEMOCRACY AND STATE REPRESSION

One purpose of the Lexical index of electoral democracy is descriptive: to differentiate regime-types in the world (Table 2) and to portray changes over time (Figure 1). Another use is to probe causal relationships between regime-type and other factors. As an example of this sort of work we shall explore the relationship between regime-type and state repression of personal integrity rights (Davenport and Armstrong 2004).

Democracies are expected to be less repressive than autocracies for a variety of reasons. First, a democratic framework is thought to promote tolerance. Second, low respect for human rights may be punished by the electorate at the ballot box. Finally, political participation and contestation provide an outlet for protests and secure legitimacy in the broader population, alleviating the extra-constitutional challenges that often spur violent government repression. Extant theory thus presents a strong prima facie case for political regime-type as an influence on state repression.

However, it is not clear what the precise empirical relationship might be. Extant work on the subject suggests three possible patterns. As summarized by Davenport and Armstrong (2004: 538–39): (1) "with every step toward democracy, the likelihood of state-related civil peace is enhanced"; (2) "human rights conditions are not only improved when full democracy exists but also when full autocracy is present"; or (3) "there may…be some threshold of domestic democratic peace, below which there is no effect of democracy on repression, but above which a negative influence can be found."

Our interest in this question is heuristic. We probe the empirical relationship between electoral democracy and state repression in order to demonstrate how the Lexical index may be brought to bear on a causal hypothesis where countries are the relevant units of analysis. Specifically, we wish to utilize the special qualities of the Lexical index in order to gain insight into the mechanisms at work in this (putatively) causal relationship.

To simplify things, we adopt the empirical format employed by Davenport and Armstrong (2004), with some minor modifications to update the analysis through 2004.[6] We

[6] Apart from Lexical, all data used are taken from the QoG standard dataset (Teorell et al. 2013).

readily grant that there are other approaches to causal modeling that might be adopted in this instance. However, since our purpose is to compare extant indices—rather than to make causal claims—difficult choices among estimators, specifications, and samples may be put aside.

Following Davenport and Armstrong, state repression is measured by the Political Terror Scale (PTS), in turn based on the State Department human rights country reports (Wood and Gibney 2010). We enlist OLS regression to assess the model and employ a battery of covariates including interstate armed conflict (UCDP/PRIO), internal armed conflict (UCDP/PRIO), military dictatorship (Cheibub, Gandhi, and Vreeland 2010), population (ln) (PWT), GDP/cap (ln) (PWT), and a one-period lag of the outcome. Democracy is measured in the first instance by the 10-point Polity Democracy index (scaled from 0 to 10) drawn from the Polity IV dataset (Marshall, Gurr, and Jaggers 2013).

Our second measure of democracy is the Lexical index, with one notable coding change. Data on state repression are available only from 1976, meaning that there is little variation in suffrage laws during the observed period. Distinctions across L4–L6 of the Lexical index are therefore rendered moot, prompting us to collapse them into a single category (L4). The resulting index has five levels—L0–L4, with roughly equal membership—and is otherwise identical to the index described above.

To probe possible causal mechanisms lying between regime-type and state repression, we adopt a series of approaches, summarized in Table 7. First, we test the possibility of a linear relationship. Polity (Model 1) and Lexical (Model 2) both indicate a negative relationship: more democracy is correlated with less repression, corroborating the general theory but leaving the problem of causal explanation opaque. Next, we test the possibility of a curvilinear relationship by introducing a multiplicative term. The coefficients for Polity (Model 3) and Lexical (Model 4) are similar, though only Polity offers support for the notion that democracy's impact on repression is nonlinear.

Finally, we attempt to explore each category of these indices separately through the use of dummy variables representing each level (with the first level omitted as a reference category). Results, shown in Models 5 and 6, are again broadly similar across the two indices, though there are some important differences. The coefficient for L1 in the Polity index is significantly more repressive than the reference category, L2–6 do not show results are statistically distinguishable

from the null, and L7–10 show negative, and statistically significant coefficients. By contrast, L1, L2, and L4 in the Lexical index are statistically significant from the reference category, but not L3.

TABLE 7

ELECTORAL DEMOCRACY AS A PREDICTOR OF STATE REPRESSION: LEXICAL AND POLITY COMPARED

	Linear		**Curvilinear**		**Disaggregated**	
	1	**2**	**3**	**4**	**5**	**6**
Polity	-.025 (.003) ***		.051 (.011) ***			
Polity^2			-.009 (.001) ***			
Polity, L1					.120 (.045) **	
Polity, L2					-.067 (.050)	
Polity, L3					-.010 (.067)	
Polity, L4					.073 (.067)	
Polity, L5					-.033 (.058)	
Polity, L6					-.011 (.040)	
Polity, L7					-.064 (.045)	
Polity, L8					-.119 (.037) ***	
Polity, L9					-.170 (.041) ***	
Polity, L10					-.400 (.037) ***	
Lexical		-.038 (.007) ***		.001 (.029)		
Lexical^2				-.009 (.007)		
Lexical, L1						-.121 (.032) ***
Lexical, L2						-.125 (.060) *
Lexical, L3						-.046 (.033)
Lexical, L4						-.220 (.031) ***
$PTSsd_{t-1}$	.691 (.011) ***	.715 (.010) ***	.669 (.011) ***	.714 (.010) ***	.659(.012) ***	.703 (.010) ***
Interstate conflict	.057 (.054)	.018 (.053)	.086 (.053)	.021 (.053)	.088(.053) *	.029 (.053)
Internal conflict	.371 (.026) ***	.375 (.025) ***	.387 (.026) ***	.377 (.025) ***	.388(.026) ***	.382 (.025) ***
Military dictator	.042 (.027)	.056 (.026) *	.096 (.028) ***	.055 (.026) *	.086(.028) **	.042 (.026)
Population (ln)	.051 (.007) ***	.043 (.006) ***	.055 (.007) ***	.043 (.006) ***	.056(.007) ***	.044 (.006) ***
GDP/cap. (ln)	-.051 (.009) ***	-.073 (.008) ***	-.025 (.010) **	-.070 (.008)***	-.022(.010) *	-.070 (.008) ***
Constant	.727 (.095) ***	.930 (.090) ***	.497 (.100) ***	.893(.094) ***	.499(.101) ***	.967 (.095) ***
R^2	.748	.770	.752	.770	.754	.772
N	3553	3924	3553	3924	3553	3924

Sample period: 1976–2004. Countries: 155/165. Estimator: OLS. Standard errors in parentheses. *<.1, **<.01, ***<.001 (two-tailed test). L=levels on an ordinal scale (not lags).

Leaving aside for a moment the question of which index offers a truer representation of the relationship between democracy and repression, let us consider what might be learned from Models 5 and 6. Davenport and Armstrong (2004: 548) conclude that "there are important differences between the political systems associated with the highest levels of the Polity measure…"—a reasonable conjecture. But they cannot follow this statement up with any speculation about what is distinctive about the higher levels of the Polity index or what might be driving the apparently curvilinear relationship between democracy and repression. This is because the levels of the Polity index are not individually interpretable. In this respect, ordinal indices of democracy such as Polity, PR, and CL function very much like interval indices. They inform us about quantities (more or less of some latent trait) but not about qualities (categorical differences across levels).

By contrast, the Lexical index provides ample fodder for theorizing because each level defines a discrete category and each category is plausibly approached as a regime-type. Let us begin by reviewing the information contained in Model 6. No level in the Lexical index reveals higher levels of state repression than level L0 (no elections). While it is unsurprising to discover that a non-electoral state has high levels of repression (for all the reasons set forth in our initial theory), it is somewhat surprising to find that there is no (statistically significant) difference in levels of repression across L0 and L3. If the model is correct, repression decreases significantly when a polity moves (hypothetically) from no national (L0) elections to a situation of national elections (L1), national multiparty elections for the legislature (L2), or—most effectively—minimally competitive elections for the legislature and the executive, while the degree of state repression in a situation of multiparty elections for legislative and executive offices that are not minimally competitive (L3) is not significantly different from a situation without national elections.

An explanation may be found in the hybrid nature of the L3 regimes, which are characterized by many of the constitutional features of democracy without the crucial missing step in which elections are allowed to become competitive. That is, L3 polities look as if they are democratic—and undoubtedly are portrayed by their leaders as democratic. But even though opposition groups are free to organize and to participate in the political system, they are not allowed to win government power (Schedler 2002). Some of the hybrid features of this setting are likely to engender more repression than in the other settings characterized by national

elections. Because the opposition is free to organize, it is likely to pose a significant challenge to the government. And because the elections are not free, the opposition is likely to pursue extra-constitutional measures, which in turn are likely to provoke government repression. In short, both government and opposition have means and motive to engage in a cycle of protest and retaliation, a setting that is likely to feed levels of state repression that are indistinguishable from settings without national elections.

EXTENSIONS

The short explanatory sketch offered in the previous section is not intended to convince. In order to be fully convincing, a causal argument would need to be accompanied by a much longer theoretical discussion intended to make sense of case-based evidence and extant theorizing on this well-trodden subject—not to mention a battery of robustness tests. Our purpose is illustrative. We hope to have shown that a lexical approach to measurement provides a useful tool for gaining insight into causal relationships and specifically into the causal mechanisms that may be at work. This feature is a product of the fact that the levels of a lexical scale are individually meaningful.

By contrast, binary scales are generally too diffuse to be useful in this context; country-years scored as 0 are different from country-years scored as 1 in many ways. It is not clear which of these differentiating features might be responsible for a causal effect or whether their impact is combinatorial. Extant ordinal scales can, in principle, be disaggregated into their component parts, as we have done with the Polity Democracy index. However, since these components are not uniquely defined, they are not very informative. We know that L3 is higher than L2, but this is about all we know. Interval scales may also be disaggregated. However, establishing the break-points is a highly arbitrary affair, and the resulting categories contain no useful information.

We suspect that the same aspects that render the Lexical index useful in the context of state repression might also be useful in the context of other research questions where regime-type lies on the right side of a causal model.

Consider the vaunted democratic peace hypothesis (Brown, Lynn-Jones, and Miller 1996). While a new scale of democracy will assuredly not solve this obdurate research question by itself, it does allow a more nuanced test of the thesis (at least as pertains to the electoral components of democracy). Specifically, we can explore whether there is a specific level in the

Lexical index beyond which conflict between nations ceases to occur and whether one or both members of the dyad must surpass that threshold. This is arguably more informative than a binary or ordinal/interval analysis of the problem.

As a second example, one might consider the contested relationship between development and democracy (Przeworski et al. 2000). With democracy on the left side of the model, one may investigate whether the empirical relationship of socioeconomic development to electoral democracy is different at various points in the Lexical index. Do increases in per capita GDP have a greater impact on electoral democracy at certain thresholds? With democracy on the right side of the model, one might investigate whether different thresholds of electoral democracy have varying relationships to economic growth (Gerring et al. 2014). For example, does the initial transition to multiparty elections have a different impact on growth performance than the transition to competitive elections?

Note that the utility of a lexical definition of democracy (like that of all others) rests ultimately on how well it explains the world around us. The electoral interpretation of democracy presumes that one dimension of democracy—grounded in elections—has the greatest impact on governance, wellbeing, and on other aspects of democracy (liberal, deliberative, et al.). It treats the electoral component as causally exogenous. Likewise, our Lexical index is premised on an idea of which features of electoral democracy are likely to be most fundamental. On this basis, we included some attributes and excluded others and arrived at a lexical ordering of those that were included. Whether this construction of the world is fruitful rests on empirical investigations that unfold over time. Our attempt in this study is to demonstrate that this approach to conceptualization and measurement bears further exploration.

It should be clear that in launching a new index of electoral democracy we are not proposing that the Lexical index has any claim to ontological priority over other sorts of indices, each of which represent certain aspects of reality and each of which has its uses. Sometimes, relationships are continuous and hence are best measured with an interval scale. Sometimes, they have only one threshold and hence are best measured with a binary scale. Our claim is that, sometimes, descriptive and causal relationships are ordinal in character or require an ordinal scale to test various threshold possibilities. In these settings, which may apply to many theories about democratic development (as cause or effect), a lexical scale may be appropriate. Here, ordinal levels are constructed in order to represent qualitatively different categories. These

categories are informative insofar as they are fecund, attaining the desiderata of any classificatory scheme, that is, to group phenomena in categories that are mutually exclusive and exhaustive.

APPENDIX A

TABLE A1

VARIABLES, DEFINITIONS, SOURCES

DEMOCRACY INDICES

BMR. To qualify as democratic a country must satisfy the following conditions: "(1) The executive is directly or indirectly elected in popular elections and is responsible either directly to voters or to a legislature, (2) the legislature (or the executive if elected directly) is chosen in free and fair elections, (3) a majority of adult men has the right to vote" (Boix, Miller, and Rosato 2013: 1531).

CL. Civil Liberties, an index measuring the respect for civil liberties (Freedom House 2013), reversed scale.

Contestation. An index derived from the first component of a principal components analysis including a large number of democracy indicators (Coppedge, Alvarez, and Maldonado 2008).

DD. Democracy-dictatorship index. To qualify as democratic a country must satisfy the following conditions: "(1) The chief executive must be chosen by popular election or by a body that was itself popularly elected, (2) The legislature must be popularly elected, (3) There must be more than one party competing in the elections, (4) An alternation in power under electoral rules identical to the ones that brought the incumbent to office must have taken place" (Cheibub, Gandhi, and Vreeland 2010: 69).

Democracy index (Polity). The upper half of the Polity2 index, stretching from 0 to 10 (Marshall, Gurr, and Jaggers 2013).

Democracy index (Vanhanen). The product of (1) the vote-share or seat-share of all but the largest party and (2) the share of adult population that voted (Vanhanen 2000).

Inclusiveness. An index derived from the second component of a principal components analysis including a large number of democracy indices (Coppedge, Alvarez, and Maldonado 2008).

Lexical. Lexical index of electoral democracy, as described in text.

PR. Political Rights, an index measuring the extent of political rights (Freedom House 2013), reversed scale.

Polity2. Polity2 index, combining Autocracy and Democracy variables, from the Polity IV dataset (Marshall, Gurr, and Jaggers 2013).

UDS. Unified Democracy Score, derived from an IRT model including a large number of democracy indicators (Pemstein, Meserve, and Melton 2010).

OTHER VARIABLES

PTSsd. Political Terror Scale (US State Department), an index measuring levels of political violence that a country experiences in a particular year based on a "terror scale" developed by Freedom House. The scale ranges from 1 ("Countries are under secure rule of law, people are not imprisoned for their views, and torture is rare or exceptional. Political murders are extremely rare") to 5 ("Terror has expanded to the whole population. The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals").

Interstate conflict. An armed conflict—as defined by the UCDP/PRIO Armed Conflict Dataset (i.e., a contested incompatibility that concerns government or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths)—that occurs between two or more states.

Internal conflict. An armed conflict—as defined by the UCDP/PRIO Armed Conflict Dataset—that occurs between the government of a state and one or more internal opposition group(s). This category also includes "internationalized internal conflicts," incompatibilities with intervention from other states on one or both sides.

Military dictatorship. A binary variable indicating whether a country is a military dictatorship, defined as a regime in which the executive relies on the armed forces to come to and stay in power (Cheibub, Gandhi, and Vreeland 2010).

Population size. Population size in thousands, from the Penn World Tables.

GDP/cap. Real GDP per capita (chain series) in constant prices, from the Penn World Tables.

TABLE A2

DESCRIPTIVE STATISTICS FOR VARIABLES USED IN TABLE 7

	Obs	Mean	SD	Min	Max
Democracy index (Polity)	4176	4.097	4.169	0	10
Lexical	5101	2.589	1.573	0	4
PTSsd	4504	2.331	1.152	1	5
Interstate conflict	4606	0.030	0.172	0	1
Internal conflict	4606	0.220	0.414	0	1
Military dictatorship	5037	0.184	0.388	0	1
Population size (ln)	4790	8.509	2.006	2.851	14.071
GDP/cap. (ln)	4728	8.327	1.295	5.081	11.158

Sample period: 1976–2004 (corresponding to analysis in Table 7).

REFERENCES

Alvarez, R. Michael, José A. Cheibub, Fernando Limongi, and Adam Przeworski. 1996. "Classifying Political Regimes." *Studies in Comparative International Development* 31:2, 3–36.

Armstrong II, David A. 2011. "Stability and Change in the Freedom House Political Rights and Civil Liberties Measures." *Journal of Peace Research* 48:5, 653–62.

Bernhard, Michael, Timothy Nordstrom, and Christopher Reenock. 2001. "Economic Performance, Institutional Intermediation, and Democratic Survival." *Journal of Politics* 63:3, 775–803.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2013. "A Complete Dataset of Political Regimes, 1800–2007." *Comparative Political Studies* 46:12, 1523–54.

Brown, Michael E., Sean M. Lynn-Jones, and Steven E. Miller, eds. 1996. *Debating the Democratic Peace*. Cambridge: MIT Press.

Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143: 1–2, 67–101.

Collier, David, and Robert Adcock. 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices about Measures." *Annual Review of Political Science* 2: 537–65.

Coppedge, Michael, Angel Alvarez, and Claudia Maldonado. 2008. "Two Persistent Dimensions of Democracy: Contestation and Inclusiveness." *Journal of Politics* 70:3, 335–50.

Coppedge, Michael, and John Gerring, with David Altman, Michael Bernhard, Steven Fish, Allen Hicken, Matthew Kroenig, Staffan I. Lindberg, Kelly McMann, Pamela Paxton, Holli A. Semetko, Svend-Erik Skaaning, Jeffrey Staton, and Jan Teorell. 2011. "Conceptualizing and Measuring Democracy: A New Approach." *Perspectives on Politics* 9:1, 247–67.

Correlates of War. 2011. *State System Membership* (v2011). http://www.correlatesofwar.org/COW2%20Data/SystemMembership/2011/System2011.html.

Davenport, Christian, and David Armstrong. 2004. "Democracy and the Violation of Human Rights: A Statistical Analysis from 1976–1996." *American Journal of Political Science* 48:3, 538–54.

Elkins, Zachary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations." *American Journal of Political Science* 44:2, 287–94.

Freedom House. 2013. *Freedom in the World 2013*. http://www.freedomhouse.org/report/freedom-world-2013/methodology, accessed December 16, 2013.

Gerring, John, Svend-Erik Skaaning, Matthew Maguire, and Eitan Tzelgov. 2014. "Democracy and Growth: A Lexical Approach." Unpublished manuscript, Department of Political Science, Boston University.

Gerring, John, Svend-Erik Skaaning, and Daniel Pemstein. 2014. "A Concept-Driven Approach to Measurement: The Lexical Scale." Unpublished manuscript, Department of Political Science, Boston University.

Gleditsch, Kristian. 2013. *List of Independent States*. http://privatewww.essex.ac.uk/~ksg/statelist.html.

Gleditsch, Kristian S., and Michael D. Ward. 1997. "Double Take: A Re-examination of Democracy and Autocracy in Modern Polities." *Journal of Conflict Resolution* 41:3, 361–83.

Goertz, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press.

Hadenius, Axel, and Jan Teorell. 2007. "Pathways from Authoritarianism." *Journal of Democracy* 18:1, 143–56.

Held, David. 2006. *Models of Democracy,* 3d ed. Stanford: Stanford University Press.

Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.

Knutsen, Carl Henrik. 2010. "Measuring Effective Democracy." *International Political Science Review* 31:2, 109–28.

Kurtz, Marcus, and Andrew Schrank. 2007. "Growth and Governance: Models, Measures, and Mechanisms." *Journal of Politics* 69:2, 538–54.

Levitsky, Steven, and Lucan Way. 2002. "The Rise of Competitive Authoritarianism." *Journal of Democracy* 13:2, 51–55.

Lively, Jack. 1975. *Democracy*. New York: St. Martin's.

Marshall, Monty G., Ted Gurr, and Keith Jaggers. 2013. "Polity IV Project: Dataset Users' Manual." http://www.systemicpeace.org/inscr/p4manualv2012.pdf, accessed December 16, 2013.

Merkel, Wolfgang. 2004. "Embedded and Defective Democracies." *Democratization* 11:1, 33–58.

Møller, Jørgen, and Svend-Erik Skaaning. 2013. "Regime Types and Democratic Sequencing." *Journal of Democracy* 24:1, 142–56.

Munck, Gerardo L. 2009. *Measuring Democracy: A Bridge between Scholarship and Politics*. Baltimore: John Hopkins University Press.

Naess, Arne, Jens Christophersen, and Stein Kvalo. 1956. *Democracy, Ideology, and Objectivity*. Oslo: Universitetsforlaget.

Pemstein, Daniel, Stephen Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18:4, 426–49.

Persson, Torsten, and Guido Tabellini. 2006. "Democracy and Development: The Devil in the Details." *American Economic Review* 96:2, 319–24.

Przeworski, Adam, Michael Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Material Well-Being in the World, 1950–1990*. Cambridge: Cambridge University Press.

Przeworski, Adam, et al. 2013. "Political Institutions and Political Events (PIPE) Data Set." Department of Politics, New York University.

Ravallion, Martin. 2011. "Mashup Indices of Development." *World Bank Research Observer* 27: 1–32.

Schedler, Andreas. 2002. "The Menu of Manipulation." *Journal of Democracy* 13:2, 36–50.

———. 2012. "Judgment and Measurement in Political Science." *Perspectives on Politics* 10:1, 21–36.

Schumpeter, Joseph A. 1950. *Capitalism, Socialism and Democracy.* New York: Harper & Bros.

Seawright, Jason, and David Collier. 2014. "Rival Strategies of Validation: Tools for Evaluating Measures of Democracy." *Comparative Political Studies* 47:1, 111–38.

Teorell, Jan, Nicholas Charron, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Petrus Sundin, and Richard Svensson. 2013. *The Quality of Government Dataset,* version 20Dec13. University of Gothenburg: The Quality of Government Institute. http://www.qog.pol.gu.se/, accessed February 1, 2014.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52:1, 201–17.

Vanhanen, Tatu. 2000. "A New Dataset for Measuring Democracy, 1810–1998." *Journal of Peace Research* 37:2, 251–65.

Vermillion, Jim. 2006. "Problems in the Measurement of Democracy." *Democracy at Large* 3:1, 26–30.

Wood, Reed, and Marc Gibney. 2010. “The Political Terror Scale (PTS): A Re-introduction and a Comparison to CIRI.” *Human Rights Quarterly* 32:2, 367–400.