VARIETIES OF DEMOCRACY

A New Approach to Conceptualizing and Measuring Democracy

Paper presented by Michael Coppedge and Jan Teorell at the 3rd International Conference on Democracy as Idea and Practice," University of Oslo, Norway, 12-13 January 2012, on behalf of:

Principal Investigators:
Michael Coppedge – University of Notre Dame
John Gerring – Boston University
Staffan I. Lindberg – University of Gothenburg & University of Florida
Jan Teorell – University of Lund

with:
David Altman – Pontificia Universidad Católica de Chile
Michael Bernhard – University of Florida
M. Steven Fish – University of California, Berkeley
Allen Hicken – University of Michigan
Matthew Kroenig – Georgetown University
Drew Linzer – Emory University
Kelly McMann – Case Western Reserve University
Pamela Paxton – University of Texas, Austin
Holli Semetko – Emory University
Svend-Erik Skaaning – Aarhus University, Denmark
Jeffrey Staton – Emory University

Note: This project description integrates material previously published in Coppedge, Gerring, et al. (2011)

Initial funding provided by…

In the wake of the Cold War, democracy has gained the status of a mantra. Yet there is no consensus about how to conceptualize and measure it well enough to support meaningful and accurate comparisons through time and across countries. Skeptics may question whether such comparisons are possible.[1] Nevertheless, there is a clear need for democracy measurement in order to mark progress and setbacks, to explain democracy, to reveal its consequences, and to affect its future course. The need is so strong that crossnational indices such as Freedom House and Polity are used all over the globe even though they are known to suffer from major shortcomings.

For policymakers, activists, academics, and citizens around the world the conceptualization and measurement of democracy matters. Billions of dollars in foreign aid intended to promote democracy and governance in the developing world is contingent upon judgments about how democratic a polity is at the present time, its recent history, future prospects, and the likely causal effects of giving or withholding assistance. Likewise, a large body of work in social science deals with these same issues, i.e., the nature, causes, consequences, and trajectories of democracy around the world.

How, then, can this task be handled most effectively? In this document we introduce a new approach to conceptualization and measurement, dubbed the Varieties of Democracy (V-Dem) project. In Part I, we review the weaknesses of extant approaches. In Part II, we lay out our approach, which may be characterized as historical, multidimensional, and disaggregated. In Part III, we discuss the process of data collection. In Part IV, we describe project management. In Part V, we discuss what we have accomplished to date and what remains to be accomplished. In Part VI, we discuss possible methods of validation. In Part VII, we suggest the payoffs our approach may bring to the study of democracy.

I. Existing Democracy Indices

Critiques of democracy indices are legion.[2] We touch briefly on seven key issues of conceptualization and measurement: definition, precision, sources and coverage, coding, aggregation, and tests of validity and reliability.

Our discussion focuses on the most prominent efforts including the Political Rights, Civil Liberty, Nations in Transit, and Countries at the Crossroads indices, all sponsored by Freedom House; the Polity2 variable from the Polity IV database; a binary measure of democracy and dictatorship ("DD") constructed by Adam Przeworski and colleagues; a binary measure constructed by Michael Bernhard, Timothy Nordstrom, and Christopher Reenock ("BNR"); a multidimensional index produced by the Economist Intelligence Unit ("EIU"); and the Bertelsmann Transformation Index ("BTI") funded by the Bertelsmann Foundation.[3] Glancing reference will be made to other indices in an increasingly crowded

[1] The authors thank the uncountable people who have generously provided comments and feedback at various stages of this project.

[2] See Acuna-Alfaro 2005; Beetham 1994; BergSchlosser 2004a, 2004b; Bollen 1993; Bollen and Paxton 2000; Bowman, Lehoucq, and Mahoney 2005; Coppedge, Alvarez, and Maldonado 2008; Foweraker and Krznaric 2000; Gleditsch and Ward 1997; Lindberg 2006; McHenry 2000; Munck 2009; Munck and Verkuilen 2002; Treier and Jackman 2008; Vermillion 2006. For work focused more generally on governance indicators, see Arndt and Oman 2006; Besancon 2003; Kurtz and Schrank 2007; Sudders and Nahem 2004; Thomas 2010; USAID 1998.

[3] Freedom House employs two indices, "Political Rights" and "Civil Liberties" (sometimes they are employed in tandem, sometimes singly) each of which extends back to 1972 and covers most sovereign and semi-sovereign nations (see www.freedomhouse.org). Polity IV (Marshall and Jaggers 2007) also provides two aggregate indices, "Democracy" and "Autocracy," usually used in tandem (by subtracting one from the other), which provides the Polity2 variable. Coverage extends back to 1800 for most sovereign countries with populations greater than 500,000 (www.cidcm.umd.edu/inscr/polity). DD (Alvarez, Cheibub, Limongi, and Przeworski 1996; Cheibub, Gandhi, and Vreeland 2010) codes countries dichotomously

field, and many of the points made in the following discussion probably apply more broadly.[4] The following exercise does not purport to provide a comprehensive review; it merely calls attention to the sort of problems that are common in this field.[5]

In our discussion, an *index* will be understood as a highly aggregated, composite measure of democracy such as Polity2, while an *indicator* will be understood as a more specific, disaggregated element of democracy such as turnout.

Definition

Democracy, understood in a very general way, means rule by the people. This common denominator claims a long heritage stretching back to the Classical age (Held 2006). All usages of the term also presume sovereignty. A polity must enjoy some degree of self-government in order for democracy to be realized.

Beyond these core definitional elements there is great debate. The debate covers both descriptive and normative aspects; what political regimes are (or reasonably could be) and what they ought to be.[6] Since definitional consensus is necessary for obtaining consensus over measurement, the goal of arriving at a single universally accepted measure of democracy is impossible as long as this great debate continues.

For example, the Polity2 index rates the United States as fully democratic throughout the twentieth century and much of the nineteenth century. This is a fair conclusion if one disregards the composition of the electorate—from which women and blacks were generally excluded—in one's definition of democracy (Paxton 2000). Similar challenges could be levied against other indices that omit consideration of attributes that some regard as definitional. These omissions are particularly glaring where democracy is defined in a minimal fashion.

More encompassing conceptions of democracy may be criticized for including elements that fall far from the core meaning of the term. For example, the Political Rights index includes questions pertaining to corruption, civilian control of the police, the absence of widespread violent crime, willingness to grant political asylum, the right to buy and sell land, and the distribution of state enterprise profits (Freedom House 2007). The authors of the index would argue that it measures freedom, not democracy; nevertheless, it is frequently used as an index of democracy.

In other instances, it is the judgments embraced within an index that are problematic. For

---

(democracy/dictatorship), including most sovereign countries from 1946 to the present. BNR (2001) construct a binary measure that extends (for all sovereign nations) from 1919 to the present. The EIU index is composed of five core dimensions and sixty sub-components, which are combined into a single index of democracy (EIU 2010). Coverage extends to 167 sovereign or semisovereign nations for 2006, 2008, and 2010.

[4] E.g., Competition and Participation variables (Vanhanen 2000), the Polyarchy index (Coppedge and Reinicke 1990), Contestation and Inclusiveness indices (Coppedge, Alvarez, and Maldonado 2008), the Political Regime Change [PRC] dataset (Gasiorowski 1996; updated by Reich 2002), the Democratization Dataset (Schneider and Schmitter 2004a), Unified Democracy Scores (Pemstein, Meserve, and Melton 2010, the Democracy Barometer (Buhlmann, Merkel, and Wessels 2008), and indicators based on Altman and Pérez-Liñán 2002; Arat 1991; Bollen 1980, 2001; Bowman, Lehoucq, and Mahoney 2005; Hadenius 1992; and Moon et al. 2006.

[5] Detailed surveys can be found in Hadenius and Teorell 2005; Landman 2003; and Munck and Verkuilen 2002. See also Acuña-Alfaro 2005; Beetham 1994; Berg-Schlosser 2004a, 2004b; Bollen 1993; Bollen and Paxton 2000; Bowman, Lehoucq, and Mahoney 2005; Casper and Tufis 2003; Elkins 2000; Foweraker and Krznaric 2000; Gleditsch and Ward 1997; McHenry 2000; Munck and Verkuilen 2002; and Treier and Jackman 2008.

[6] Studies of the concept of democracy are legion; see, e.g., Beetham 1994, 1999; Collier and Levitsky 1997; Held 2006; Lively 1975; Saward 2003; Weale 1999. All emphasize the far-flung nature of the core concept.

example, the EIU index regards mandatory voting as reflecting negatively on the quality of democracy in a country. While this provision infringes upon individual rights and in this respect may be considered undemocratic, it also enhances turnout and, arguably, the quality of representation. Its status in enhancing rule by the people is therefore controversial, depending on one's conception of democracy..

The methodological problems affecting contemporary indices therefore begin at the level of definition. Our solution, discussed in more detail below, is to create indices of *all* the main varieties of democracy found in the literature, as well as their component parts, so that countries can be compared over a range of definitions.

Precision

Many of the leading democracy indices are insensitive to important gradations in the degree or quality of democracy. If one purpose of any measurement instrument is discrimination (Jackman 2008), extant democracy indices fall short of the ideal.

At the extreme, binary measures such as DD and BNR reduce democracy to a dummy variable useful for purposes such as analyzing the duration of democratic regimes (Collier and Adcock 1999). However, this dichotomous coding lumps together polities that exhibit quite different regime qualities (Elkins 2000). For example, the DD index recognizes no distinctions within the large category of countries that have competitive elections and occasional leadership turnover. Papua New Guinea and Sweden thus receive the same score, despite evident differences in the quality of elections, civil liberties, and barriers to competition.

Continuous measures appear to be more sensitive to gradations of democracy/autocracy because they have more ranks. Freedom House scores democracy on a seven-point index (13 points if the Political Rights and Civil Liberties indices are combined). Polity provides a total of 21 points if the Democracy and Autocracy scales are merged (an aggregation procedure suggested by the data providers), creating the Polity2 variable. Appearances, however, can be deceiving. Polity scores, for example, bunch up at a few thresholds (notably -7 and +10), suggesting that the scale is not as sensitive as it purports to be. Likewise, because Polity2 combines six underlying factors, a country's score is a reflection of its subscore on these six factors (and the weighting system). This means that two countries with the same score may have quite different regimes--the divergences are, in effect, averaged away. The EIU index is the most smoothly continuous of the extant indices, but like Polity, aggregate scores are reflections of underlying components that can be quite different yet ignored by averaging.[7]

In sum, the precision or reliability of all indices is too low to justify confidence that a country that scores a few points higher than another is actually more democratic (Pemstein et al. 2010; Treier and Jackman 2008). Note that most extant indices are bounded and there is no way to distinguish the quality of democracy among countries that have perfect scores. This is acceptable so long as there really is no difference in the quality of democracy among these countries. Sweden's democracy has not changed since the extension of suffrage in 1917 in Polity. In 2004, Freedom House assigned the highest score on its Political Rights index to countries as dissimilar as Andorra, Bulgaria, Denmark, Israel, Mauritius, Nauru, Panama, South Africa, Uruguay, and the United States. There are substantial differences in the quality of democracy among these diverse polities, but these differentiating elements are missing from extant democracy indices (Lijphart 1999).

[7] Questions can also be raised about whether these indices are properly regarded as interval scales; see Treier and Jackman (2008). We do not envision an easy solution to this problem although Pemstein, Meserve, and Melton (2010) offer some intriguing ideas.

Sources and Coverage

Many democracy indices are limited in temporal or country coverage. Nations in Transit covers only the post-communist states. Countries at the Crossroads covers seventy countries (beginning in 2004) deemed to be strategically important and at a critical juncture in their trajectory. The Political Rights and Civil Liberty indices stretch back to 1972 and include most sovereign and semisovereign states. DD begins in 1946, BNR begins in 1919, BTI begins in 2003, and EIU begins in 2006.

Only a few democracy indices stretch back further in historical time—notably, Polity (1800-), and the index of democratization created by Tatu Vanhanen (1810–present).[8] We suspect that the enduring value of Polity stems partly from its comprehensive historical coverage—though it excludes states with fewer than 500,000 inhabitants and makes no attempt to include colonies prior to independence, even if they enjoyed substantial self-government.

The sources employed to provide coding for extant indices are often problematic. For example, the Political Rights and Civil Liberty indices rely heavily on secondary accounts such as the *New York Times* and Keesing's Contemporary Archives for coding in the 1970s and 1980s. These historical sources, while informative, do not provide equally comprehensive coverage of every country in the world, introducing potential bias into the resulting indices. In later eras, these indices have relied much more on country expert coding. However, the change from one source of evidence to another—coupled with some possible changes in coding procedures—may have compromised the continuity of the time-series. No effort has been made to revise previous scores so that they are consistent with current coding criteria and expanded knowledge of past regimes.[9]

Some indices such as the EIU rely heavily on polling data, which is available on a non-comparable and highly irregular basis for 100 or so nation-states. For other countries (about half of the population covered by the EIU) these data must be estimated by country experts or imputed. Procedures employed for this estimation are not made publicly available.[10] Although surveys of citizens are important for ascertaining attitudes, they are not available for every country in the world, and in no country are they available on an annual basis. Moreover, use of such surveys severely limits the historical reach of any democracy index, since the origin of systematic surveying stretches back only a half-century (in the US and parts of Europe) and is much more recent in most countries. Other survey-based questions are of questionable relevance for understanding the quality of democracy in a polity. It is of course interesting to know whether citizens regard their country as democratic, whether they support democratic institutions and practices, and whether they subscribe to democratic norms such as tolerance. But it is not clear whether such attitudes make a country more or less democratic. It is striking that popular support for democracy, as gauged by standardized surveys such as the Afrobarometer and LAPOP, sometimes varies inversely with standard indices of democracy (such as Polity2 and the Political Rights index) that rest on the institutional features of polities.

Expert Coding

Several indices (including Freedom House, EIU, and BTI) rely on expert coding. Such judgments can be

---

[8] Vanhanen 2000; Boix and Rosato 2001 extend the coding procedures laid out by DD back to 1800. However, we are aware of only one study employing this dataset (Boix and Stokes 2003), and the dataset has not been formally published anywhere, or updated (to the best of our knowledge), so we do not include it in the discussion.

[9] Gerardo Munck, personal communication, 2010.

[10] Reliance on survey data also raises even more difficult questions about validity, i.e., whether the indicators measure what they are supposed to measure. There is surprisingly little empirical support for the notion that respondents are able to assess their own regimes in a cross-nationally comparable way or that they tend to live under regimes that are congruent with their own values.

made fairly reliably if there are clear and concrete coding criteria (Munck and Verkuilen 2002). Unfortunately, this is not always the case. For example, the Nations in Transit expert survey poses five sub-questions to answering the question, "Is the country's governmental system democratic?":

1. Does the Constitution or other national legislation enshrine the principles of democratic government?
2. Is the government open to meaningful citizen participation in political processes and decision-making in practice?
3. Is there an effective system of checks and balances among legislative, executive, and judicial authority?
4. Does a freedom of information act or similar legislation ensure access to government information by citizens and the media?
5. Is the economy free of government domination?[11]

Quite aside from the debatable validity of equating democracy with separation of powers and a free-market economy, these are not easy questions to answer, and their difficulty stems from the general, vague, or ambiguous terms in which they are posed. One cannot judge whether the "principles of democratic government" are "enshrined" without specifying what those principles are. What does it mean to be "open to meaningful citizen participation"? What is the basis for determining whether checks and balances are "effective"? What degrees and kinds of government regulation and ownership can be permitted before economic "freedom" is infringed upon? Wherever the questions do not define these criteria, respondents must rely on their own hidden beliefs and assumptions. This creates a danger that coding decisions about *particular* topics—e.g., press freedom—will reflect an overall sense of how democratic a country is rather than an independent evaluation of the level of press freedom. In this respect, "disaggregated" indices may actually be considerably less disaggregated than they appear. It is the ambiguity of the questionnaires underlying these surveys that foster this sort of premature aggregation.

Aggregation

Since democracy is a multi-faceted concept all composite indices must wrestle with the aggregation problem— which indicators to combine into a single index, whether to add or multiply them, and how much to weight them. It goes without saying that different solutions to the aggregation problem lead to quite different results (Munck and Verkuilen 2002; Munck 2009; Goertz 2006). This is a very consequential decision, and because it is especially central to the V-Dem project we shall spend a bit of time discussing it.

The aggregation rules used by almost all existing measure are additive, with an (implicit or explicit) weighting scheme; they are sums, averages, or weighted averages of various components. It is far from obvious that this is the most appropriate aggregation rule. Others recommend that one should consider the various sub-components of democracy as necessary (non-substitutable) or mutually constitutive (interactive) (Goertz 2006: 95–127; Munck 2009; Schneider 2010). In order for any aggregation scheme to be successful, rules must be clear, they must be operational, and they must reflect an accepted definition of democracy. Otherwise, the resulting measure is not valid. Although most indices have fairly explicit aggregation rules, they are sometimes difficult to comprehend and consequently to apply (e.g., Polity). They may also include "wild card" elements, allowing the coder free rein to assign a final score, in accordance with his or her overall impression of a country (e.g., Freedom House). And they may be altogether inappropriate when the relevance or weight of one component depends on the values of other components.

A more inductive approach may also be taken to the aggregation problem. Michael Coppedge, Angel Alvarez, and Claudia Maldonado (2008) apply an exploratory factor analysis of a large set of

[11] Report on Methodology, downloaded from www.freedomhouse.hu/images/nit2009/methodology.pdf.

democracy indicators, identifying two dimensions which they label Contestation and Inclusiveness. Other writers analyze extant indices as reflections of a (unidimensional) latent variable.[12] These inductive approaches allow for the incorporation of diverse data sources and may provide uncertainty estimates for each point score. However, problems of definition are implicit in any factor-analytic or latent-variable index, for an author must decide which indicators to include in the sample— requiring a judgment about which extant indices are measuring "democracy" and which are not—and how to interpret commonalities among the chosen indicators. This is not solvable simply by referring to the labels assigned to the indicators in question, as many of the most well known and widely regarded democracy indices are packaged as "rights," "liberties," or "freedom" rather than democracy, and do not necessarily measure exactly the concepts they purport to measure. Moreover, while factor-analytic and latent variable approaches allow for the incorporation of multiple sources of data, thereby reducing some sources of error, they remain biased by any systematic error common to the chosen data sources.

Another approach to the aggregation problem is to collect primary data at a disaggregated level, letting end-users decide whether and how to aggregate it. Democracy assessments (aka audits) provide detailed indicators for a single country.[13] This sort of detailed inquisition into the quality of democracy is very much in the spirit of our enterprise, except that the collected data are limited to a single country (and often to recent years) and are not comparable across other countries where detailed assessments have been conducted.

Several specific topics integral to democracy have been successfully measured on a global scale. And several broader ventures to measure democracy in a comprehensive and disaggregated fashion have been proposed, but not fully implemented (e.g., Beetham et al. 2001; Buhlmann, Merkel, and Wessels 2008). These efforts at disaggregation push in the right direction. However, they are problematic on one or several accounts: (a) the number of indicators are very small; (b) the indicators are still highly abstract and hence difficult to operationalize; (c) the underlying components, while conceptually distinct, are gathered in such a way as to compromise their empirical independence; (d) the information necessary to code the indicator is not available across nations or prior to the contemporary era; or (e) the indicators are not released to the general public.

Consider the Polity index, which is ostensibly disaggregated into five components: competitiveness of participation, regulation of participation, competitiveness of executive recruitment, openness of executive recruitment, and constraints on executive. Although each of these components is described at length in the Polity codebook (Marshall and Jaggers 2007), it is difficult to say precisely how they would be coded in particular instances, or how the stated aggregation principles lead to an overall score for a given country in a given year (Munck and Verkuilen 2002). Even in disaggregated form (e.g. Gates et al. 2006), the Polity index is highly abstract, and therefore open to diverse interpretations.

The two principal Freedom House measures—Civil Liberties and Political Rights—are similarly difficult to interpret. Since 2006, Freedom House has released coding scores for the components of Civil Liberties and Political Rights. The Political Rights index is shown to be the product of (a) Electoral Process, (b) Pluralism and Participation, and (c) Functioning of Government. The Civil Liberties index comprises (a) Freedom of Expression, (b) Association and Organizational Rights, (c) Rule of Law, and (d) Personal Autonomy and Individual Rights. This represents a step towards disaggregation, yet intercorrelations among the seven components are extremely high—Pearson's r = 0.86 or higher. This by itself is not necessarily problematic; it is possible that all democratic (or nondemocratic) things go together. However, the high inter-correlations of the Freedom House indicators coupled with their ambiguous coding procedures suggest that these components may not be entirely independent of one

[12] Pemstein, Meserve, and Melton (2010); Bollen and Jackman (1989); Bollen and Paxton (2000); Treier and Jackman (2008).

[13] E.g., Beetham 1994, 2004; Diamond and Morlino 2005; Landman and Carvalho 2008; Proyecto Estado de la Nación 2001.

another. It is impossible to discard the possibility that country coders have a general idea of how democratic each country is, and that this idea is reflected in consistent scores across the multiple indicators. Components that are scored separately on a questionnaire may not be independently coded.

The EIU (2010) index does a better job of disaggregating components, which are reported for five dimensions: electoral process and pluralism, civil liberties, the functioning of government, political participation, and political culture. Correlations are still quite high, ranging from .74 to .93 (except for the cultural variable, which is distinctive). Moreover, the specificity of the questions makes the claim of independence among these five variables plausible. Unfortunately, EIU is unwilling to divulge data for the sixty specific questions that compose the five dimensions, so it is difficult to judge the accuracy and independence of the index. It may be useful, or it may be not, but we cannot know for sure.

Validity and Reliability Tests

Worries about validity in extant democracy indices are nourished by periodic appraisals focused on specific countries. A recent study by scholars of Central America alleges major flaws in coding for Costa Rica, El Salvador, Honduras, Guatemala, and Nicaragua in three crossnational indices—generated by Polity, Vanhanen, and Gasiorowski (1996) — errors that, the authors suspect, also characterize other indices and other countries (Bowman, Lehoucq, and Mahoney 2005).

Surprisingly, inter-coder reliability tests are not common practice among democracy indices. Freedom House does not conduct such tests, or at least does not make them public. Polity used to do so, but it required a good deal of hands-on training before coders reached an acceptable level of coding accuracy. This suggests that other coders might not reach the same decisions simply by reading Polity's coding manual. And this, in turn, can contribute to the problem of conceptual validity, in which key concepts are not well matched to the empirical data.

These critiques notwithstanding, defenders of contemporary indices often point out that extant indices are highly intercorrelated. Indeed, the correlation between Polity2 and Political Rights – the dominant indices, by most accounts – is a respectable 0.88 (Pearson's r). Yet, closer examination reveals that the consensus is largely the product of countries lying at the democratic extreme—Canada, Sweden, the United States, *et al*. When countries with the top two scores on the Freedom House Political Rights scale are eliminated, Pearson's r drops to 0.63. This is not an impressive level, especially when one considers that scholars and policymakers are usually interested in precisely those countries lying in the middle and bottom of the distribution—countries that are undemocratic or imperfectly democratic.[14]

These measurment differences are part of the reason for some of the divergent findings in empirical work where democracy is a key variable. Przeworski and Limongi (1997), for example, found that per capita income was not associated with transitions to democracy, but Zachary Elkins (2000) showed that their result depended on using a binary measure of democracy; when graded measures such as Freedom House or Polity were substituted, the correlation between income and transitions returned to significance. More generally, Gretchen Casper and Claudiu Tufis (2003) showed that few explanatory variables (beyond per capita income) have a consistently significant correlation with levels of democracy when different democracy indices are employed. In fact, the *only* predictor that both remained significant regardless of how democracy was measured and survived all their robustness checks was Rae's index of party-system fractionalization (which measures .[15] Thus, we have good reasons to suspect that extant indices suffer problems of validity and reliability and that these problems

[14] For extensive cross-country tests see Hadenius and Teorell 2005.

[15] See also Hadenius and Teorell 2005.

are consequential. They impact what we think is going on in the world.

II. A New Approach

The task of constructing a global index of democracy that is valid, precise, and universally accepted is practically impossible, for all the reasons we have discussed. If all we needed to do were to identify major regime changes, gross differences in levels of democracy, or track trends in the average global level of democracy, we could probably get by with existing measures. But development agencies, international organizations, NGOs, journalists, educators, and researchers need measures that can do more that this. Existing measures of democracy are especially inadequate for measuring smaller changes and differences in the degree of democracy or autocracy; empirically analyzing relationships among various elements of democracy; and evaluating the effectiveness of targeted democracy promotion efforts. Polity, Freedom House, and their counterparts are inadequate for these tasks. Just as critically, none of them provides a variety of measures that would enable scholars, institutions, policy-makers, and civil society actors to measure exactly what they mean by "democracy" so that they can support their comparisons and contrasts with reliable empirical data.

Four features, considered together, distinguish our approach to conceptualizing and measuring democracy. First, we take a multi-variety approach to conceptualizing democracy providing aggregate measures of all six major varieties of democracy, as well as their components. Second, we collect information on indicators relevant to democracy at a highly disaggregated level. Third, we will offer multiple aggregation techniques for all higher-order concepts, each with clear and transparent rules that can easily be replicated by end-users. Finally, we extend indicators of democracy back through modern history to 1900, wherever possible.

Before introducing these features we must further clarify the scope conditions of the project. Our principal concern is with the operation of democracy within large and fairly well-defined political units (e.g., nation-states) which we shall refer to as polities. The sizeable population of these units requires representative institutions (insofar as they are democratic), although it certainly does not preclude more direct forms of citizen governance existing side-by-side with representative institutions. We are less concerned with democracy within very small communities (e.g., neighborhoods, school boards, corporations), in contexts where the political community is vaguely defined (e.g., transnational movements), or on a global level (e.g., the United Nations).[16]

Finally, it is important to state that the "democracy" is not coterminous with "governance." To be sure, many elements of democracy might also be considered elements of good governance. Others probably have little relevance for the quality of governance and some may have a negative impact. Thus, to say that an indicator or component provides a measure of democracy is not to say that it advances the cause of good governance – or, for that matter, of justice or the good (Arneson 2004). The relationship of democracy to other desired outcomes is an empirical matter, one on which we anticipate V-Dem will be able to shed light. (See section VI.) However, measures of good governance (unless they happen to dovetail with measures of democracy) must be gathered from other sources.

History

While policymakers are rightly concerned the course of future events and most democracy scholars

[16] This is not to say that the concept of democracy should be restricted to formal and well-defined polities. It is simply to clarify our approach, and to acknowledge that different strategies of conceptualization and measurement may be required for different subject areas.

analyze contemporary issues, understanding both the present and the future requires a rigorous analysis of the past. The political regimes that exist today, and those that will emerge tomorrow, are the products of many complex processes that unfold over decades. Regime changes are sometimes sudden but, like earthquakes, these dramatic events are best understood as a combination of pent-up forces that build up over long spans of time and the precipitating factors that release them. Democratization trends, causes, effects, and sequences cannot be understood without long time series data.[17]

The advantage of our topic (in contrast with other historical measurement tasks such as national income accounts) is that much of the evidence needed to code features of democracy is well preserved in books, articles, newspapers archives, and memories. Democracy is a high-profile phenomenon. Although a secretive regime may hide the true value of goods and services, it cannot disguise the existence of an election. The features of an election that might prejudice the outcome towards the incumbent are difficult to obscure completely. This, again, is because virtually everyone living in that country, studying that country, or covering that country for some foreign news organization or aid organization has an interest in tracking this result.

Thus, we regard the goal of historical data-gathering as essential and also realistic, even if it cannot be implemented for every indicator that is gathered for the present era. Some historical indicators are better than none at all. Furthermore, if V-Dem can demonstrate empirically that the kind of indicators that are available for the past are highly correlated with "better" indicators that be produced only for recent years, cost-saving proxy indicators can be constructed. Finally, V-Dem's pilot phase proved that it is possible to extend our detailed, disaggregated indicators far back in modern history, even for little-studied countries (Suriname, Myanmar, Albania) and even during colonial years. V-Dem therefore aims to gather data whenever possible back to 1900 for all countries and colonies that can claim a sovereign or semi-sovereign existence. This means that Ghana is coded from 1914-1956 as the Gold Coast colony, British Togoland (1914-57), and the Ashanti and Fante protectorates--dependencies of Great Britain--and from 1957 as a sovereign state.[18]

V-Dem will provide codings reflecting historical changes as precisely as possible. The coding procedures allow coders to specify the exact dates (day/month/year) of changes to a particular institution, rather than lumping all events together within a given year. Likewise, elections are coded as events happening on a specific date rather than as an indefinite period of time. Naturally, date-specific data can be aggregated at 12-month intervals, which may be essential for time-series where country/years form the relevant units of analysis.

Multiple Principles

There is no consensus on what democracy at large means, beyond the prosaic notion of rule by the people. Political theorists have been emphasizing this point for some time, and empiricists would do well to take the lesson to heart (Gallie 1956; Held 2006; Shapiro 2003: 10–34).

Even so, there appears to be relative consensus on the various normative principles that can constitute this protean term. Seven key principles seem paramount: electoral, liberal, majoritarian, consensual, participatory, deliberative, and egalitarian, as summarized in Table 1. Each represents a different way of understanding what is required for "rule by the people." Thus, while no single principle can reasonably purport to embody all the meanings of democracy, these seven principles, taken

[17] This echoes a persistent theme of historically grounded social science (Nunn 2009; Mahoney and Rueschemeyer 2003; Steinmo, Thelen, and Longstreth 1992).

[18] We do not attempt to code this territory before 1914, when it was German Togoland.

together, offer a fairly comprehensive accounting of the concept of democracy as it is employed today.[19]

The *electoral* principle of democracy—identified with contestation, competition, and elite minimal, realist, or Schumpeterian democracy—is the idea that democracy is achieved through competition among leadership groups, which vie for the electorate's approval during periodic elections before a broad electorate. Parties and elections are the critical instruments in this largely procedural account of the democratic process. Of course, many additional factors might be regarded as important for ensuring and enhancing electoral contestation, e.g., civil liberties, an active media, a written constitution, an independent judiciary (to enforce the rules of the game), and so forth. However, these factors are viewed as secondary to electoral institutions (See Dahl 1956; Przeworski et al. 2000; Schumpeter 1950). A minimal degree of electoral democracy is a precondition for the other principles of democracy in V-Dem. That is, it makes little sense to say that there is liberal, majoritarian, participatory, deliberative, or egalitarian democracy in a polity unless key law- and policymaking bodies at the national level are elected and those elections are minimally free and fair.

The *liberal* (sometimes called pluralist) principle of democracy stresses the intrinsic importance of transparency, civil liberty, rule of law, horizontal accountability (effective checks on rulers), and minority rights. These are seen as defining features of democracy, not simply as aids to political competition. The liberal model takes a "negative" view of political power insofar as it judges the quality of democracy by the limits placed on government. Principles and procedures must be established so as to ensure that rule by the majority does not result in the oppression of minorities or the loss of individual liberties.[20]

The *majoritarian* principle of democracy (aka responsible party government) reflects the principle that the will of the majority should be sovereign. The many should prevail over the few. To facilitate this, political institutions must centralize and concentrate, rather than disperse, power (within the context of competitive elections), which means that majoritarian democracy is in tension with liberal democracy in many respects, e.g., strong and centralized parties, a unitary rather than federal constitution, plurality rather than proportional electoral laws (or PR with high statutory thresholds), and so forth. Even so, many aspects of democracy, such as civil liberties, due process, human rights, and transparency, are compatible with both liberalism and majoritarianism.[21]

Majoritarian democracy is in a different kind of tension with *consensual* democracy, which favors the inclusion of as many political perspectives as possible in decisionmaking rather than decisionmaking by narrow majorities. Consensual democracy therefore emphasizes proportional electoral laws, fragmented party systems, larger-than-necessary cabinets, supermajority decision rules, separation of powers, a federal constitution, and other institutions that require the national head of government to share power with other political actors and bodies. It also values liberal principles such as civil liberties, due process, and human rights even more than majoritarian democracy does.[22]

The *participatory* principle of democracy is usually viewed as a lineal descendant of the "direct"

---

[19] The consensus is only relative: even on our team there is disagreement about whether we should list seven, six, or five principles, whether they should be called "principles" or "dimensions," and in particular whether "majoritarian" and "consensual" should both be listed when they can be thought of as opposite poles on a single dimension. This list of seven principles is not intended to resolve these difficult issues once are for all, but only to provide some assurance that we are trying to take a comprehensive inventory of ways of thinking about democracy that is likely to include almost all the components that any user would want to have measured.

[20] See Dahl 1956 on "Madisonian Democracy"; see also Gordon 1999; Hamilton, Madison and Jay 1992; Hayek 1960; Held 2006, ch. 3; Hirst 1989; Mill 1958; Vile 1998.

[21] See American Political Science Association 1950; Bagehot 1963; Ford 1967; Goodnow 1900; Lijphart 1999; Lowell 1889; Ranney 1962; Schattschneider 1942; Wilson 1956.

[22] See Lijphart 1999.

(i.e., nonrepresentative) model of democracy, as derived from the experience of Athens—though elements of this model may also be discerned in "republican" thought and in the experience of many small communities throughout the world and throughout human history (Pocock 1975). The motivation for participatory democracy is uneasiness about delegating complete authority to representatives. Direct rule by citizens is preferred, wherever practicable. And within the context of representative government, the participatory component is regarded as the most democratic element of the polity. This model of democracy thus highlights the importance of voting, but also of citizen assemblies, party primaries, referenda, juries, social movements, public hearings, town hall meetings, and other forums of citizen engagement.[23]

The *deliberative* principle of democracy focuses on the process by which decisions are reached in a polity. A deliberative process is one in which public reasoning focused on the common good motivates political decisions—as contrasted with emotional appeals, solidary attachments, parochial interests, or coercion. In this principle, democracy requires more than a mindless aggregation of existing preferences; there should be respectful dialogue at all levels—from preference formation to final decision—among informed and competent participants who are open to persuasion (Dryzek 2010: 1). "The key objective," writes David Held (2006: 237), "is the transformation of private preferences via a process of deliberation into positions that can withstand public scrutiny and test." Some political institutions have a specifically deliberative function, such as consultative bodies (hearings, panels, assemblies, courts); polities with these sorts of institutions might be judged more deliberative than those without them. However, the more important issue is the degree of deliberativeness that can be discerned across all powerful institutions in a polity (not just those explicitly designed to serve a deliberative function) and among the citizenry.[24]

The *egalitarian* principle of democracy stresses a 'positive' understanding of political equality. It is not enough to have the formal ('negative') political rights and civil liberties. The polity should also address material and immaterial inequalities that inhibit the actual use of rights and liberties by some groups in society. These inequalities can exist with respect to class, gender, ethnicity, language, religion, caste, or other ascriptive groups. An egalitarian polity is one that not only offers but actually achieves roughly equal participation, equal representation, and equal protection. For this, equal resources in key areas such as income, education, and health are presumed to be a key feature of political empowerment. Political equality thus presumes that the polity has facilitating institutions that actively address social inequality and engage in some degree of redistribution.[25]

Table 1 summarizes these principles. It also refers to related "conceptions" of democracy. The "principles" can be thought of as relatively pure and narrow concepts that hew closely to an ideal type. That is, there may be no existing polity that realizes each principle to the exclusion of the other principles (even with the exception of electoral democracy, which is common to them all). The "conceptions," by contrast, are more fully specified or "thicker" concepts that correspond more closely to real-world polities. Conceptions, therefore, emphasize the principle for which they are named but also include components primarily associated with other principles: they overlap much more. Thus, a

---

[23] See Barber 1988; Benelo and Roussopoulos 1971; Fung and Wright 2003; Macpherson 1977; Mansbridge 1983; Pateman 1976.

[24] See Bohman 1998; Cohen 1989; Elster 1998; Fishkin 1991; Fung 2005; Gutmann and Thompson 1996; Habermas 1984, 1996; Held 2006, ch. 9. A number of recent studies have attempted to grapple with this concept empirically; see Bächtiger 2005; Dryzek 2009; Mutz 2008; Ryfe 2005; Steiner et al. 2004; Thompson 2008.

[25] See Dahl 1982, 1989; Lindblom 1977. Many of the writings cited previously under participatory democracy might also be cited here. Taking a somewhat different stand on this issue, Beetham 1999 and Saward 1998: 94-101 do not request an equal distribution of resources. Rather, they consider access to basic necessities in the form of health care, education, and social security to be democratic rights as they make participation in the political process possible and meaningful.

conception of really-existing participatory democracy should include not only the electoral principle, but also the political rights and guarantees of liberalism, a modicum of egalitarianism, and some institutions associated with either majoritarian or (more likely) consensual principles. The principles are more useful for theoretical analyses that require discriminating among ideal types; the related conceptions are likely to be more useful for describing and comparing actual polities.

Naturally, the conceptual scheme presented in Table 1 does not capture all the theoretical distinctions at play. It does not fully capture the distinction between direct and representative democracy, or among different principles of representation (Pitkin 1967). It does not capture the intellectual history or etymology of the concept, though traces of that history will be glimpsed in each principle (Dunn 1995; Shapiro and Hacker-Cordon 1999). All typologies are limited in some respects. Nevertheless, the typology summarized in Table 1 captures a good deal of the action surrounding current debates on democracy. Each principle is logically distinct and—at least for some theorists—independently valuable. For example, some writers believe that enhanced avenues for participation are good for democracy even in the absence of liberal aspects of democracy.

Moreover, we suspect that there is a good deal of divergence across these six principles among the world's polities. Some will be particularly strong on the electoral principle; others will be strong on the egalitarian principle, and so forth. Thus, the typology of ideal-types provided in Table 1 will prove a useful empirical device, allowing one to chart variation in political institutions through time and space.

Disaggregation

The six principles discussed in the previous section take one step in the direction of disaggregation. A second step is the identification of mid-level *components*. These specify and operationalize more fully different aspects of the varieties of democracy and move us closer to the very specific indicators that can be measured most validly and reliably. In some areas, components must be divided into sub-components before moving on to indicators.

The final step in disaggregating the concept of democracy is the identification of *indicators*. These are measurable empirical referents of components. In identifying indicators we look for features that connect each component to a particular principle of democracy; bring the political process into closer alignment with the core meaning of democracy (rule by the people); and are measurable—directly or indirectly (via coder judgments) across polities and through time.

Appendix A (V-Dem: List of Indicators) lists all indicators integrated into the V-Dem project.[26]

The V-Dem conceptual scheme has several levels of specificity, as follows:

Core concept (1)

Principles and Related Conceptions (7)

Components (number to be determined)

Sub-Components (number to be determined)

Indicators (399)

Appendix B contains taxonomic (strictly hierarchical) depictions of the conceptual schemes for

[26] Note that some indicators refer to *de jure* aspects of a polity – that which is stipulated by statute or constitutional law (including the unwritten constitution of places like the United Kingdom and well-established principles of constitutional law, as developed through a common law system). Others refer to *de facto* aspects of a polity – that which is true in practice. Frequently, we suspect that the latter is at variance with the former; hence, the need for including both de jure and de facto indicators in the V-Dem project.

each conception. This kind of taxonomy is useful for enumerating all the indicators by classifying them under the most relevant sub-component, component, and conception or principle. However, a strict taxonomy cannot capture the conceptual schemes fully because some indicators are relevant for measuring multiple sub-components, some sub-components are relevant for multiple components, and some components are relevant for multiple principles and, especially, conceptions. For example, the sub-component "executive selection and removal by election" is relevant for both electoral democracy and liberal democracy; civil liberties are relevant for both participatory and egalitarian democracy; and several components--judicial independence, subnational elections, number of parties in the cabinet, and so on--are relevant for both majoritarian and consensual democracy, albeit in opposite senses. These more complex relationships are best depicted with interactive software. One such visualization is available at the V-Dem site. The reader is encouraged to browse this visualization to see the relationships that we envision across conceptions, components, and indicators.

Further clarifications pertaining to the set of indicators contained in Appendix A and the taxonomy contained in Appendix B can be found below.

Aggregation Techniques

Of course, the taxonomy and visualization tell us only about inter-relationships among conceptions, components, and indicators at the level of membership – what belongs with what. We must also consider the problem of weighting, i.e., how to assign weights to the various elements of an index (a conception or component). This will be referred to as a problem of aggregation.

As part of this project, we will experiment with multiple aggregation schemes for each conception and component. Ultimately, we may decide that one or several of these aggregation techniques are superior, in which case this will be presented in the final version of V-Dem. This is a matter that can only be determined at a later stage of the project, when data collection is substantially complete. At present, it is sufficient to indicate the various approaches that will be implemented and assessed.

A first step is to decide whether each indicator is to be treated as a cause indicator, an effect indicator, or an intrinsic (defining) attribute (Goertz 2006; Bollen 1993). This distinction matters because we expect effect indicators to be intercorrelated but do not necessarily expect this of cause indicators or intrinsic attributes. For example, it makes sense to think of indicators of freedom of movement, religion, expression, and association, and freedom from servitude, torture, and murder as reflections or manifestations (effects) of an underlying dimension of civil liberties and human rights; and therefore to expect them all to be intercorrelated. In contrast it makes sense to think of life tenure, fixed budgets, and a consensual judicial appointment process as causes of the judicial independence component, whether they are intercorrelated or not, because they could each have an independent influence on the latent variable.

The second step in aggregation is to test the designated effect indicators for internal consistency. If our expectations prove false--for example, if an indicator is not significantly related to and well predicted by the latent variable, or if the measurement model does not fit the data well--we will either exclude the stray indicators from the scale or rethink how best to use them. Indicators that are not strongly correlated may be combined into a meaningful index when there is good theoretical guidance, especially about whether a high score on one indicator can compensate for a low score on another (substitutability).

Third, part of this theoretical guidance often consists of conditional relationships among indicators, based on common understandings. For example, the extent to which elections are free and fair matters little if suffrage is severely limited. Similarly, we would insist that indicators of political

equality not increase the value of an egalitarian democracy index unless certain thresholds of free and fair elections and civil liberties are passed. Aggregation formulas will reflect these conditional relationships. For example, indicators could be combined into a cumulative scale, in which the more peripheral attributes of the concept do not count unless more central prerequisites are satisfied.

Finally, aggregation requires choosing a technique. We will aggregate variables into indices in four distinct ways. First, we will follow a basic scaling approach that simply averages together a series of ratings that are theoretically associated with the underlying concepts of interest. We expect, however, that due to inherent limitations in the quantity and precision of rater codings, these scales will be overly noisy and suffer from other problems of validity and reliability. If this is determined to be the case, we will move to the application of more sophisticated latent variable models. A second option is therefore to use conventional factor or principal components analysis to assign factor weights based on correlations between each item and a latent variable. A third technique is based on item response theory (IRT) models, which are a statistical approach for aggregating discrete outcomes into continuous underlying scales in a logically coherent way. These models allow for stochastic rater error, in addition to accommodating the possibility that not every item is an equally strong indicator of the core latent concept. Finally, we are developing a novel multilevel, multi-rater latent variable model model that is specifically designed for application to data in our project.

Aggregating the data in different ways will reveal whether the more mathematically sophisticated methods are actually necessary for producing valid, reliable, and meaningful indices, as opposed to simpler additive indices that would be easier for most users to understand. We consider this an exciting new avenue for research that has not been possible with extant democracy measures.

IRT produces latent scales by estimating how a series of observed variables interrelate across a group of cases – here, countries and years. The technique estimates both a series of thresholds or "difficulty" parameters and a set of "discrimination" parameters for each item. Cases that cross only "easy" thresholds are scored low on the scale and those that can cross "difficult" thresholds are scored high. The "discrimination" of each item indicates how informative that item is for placing cases along the latent scale. If an item has low discrimination (that is, close to zero), then knowledge of that item is only minimally informative for determining a case's location on the scale. Since many of our ratings are recorded as ordinal variables with more than two possible outcomes, we will use a modification of the basic dichotomous-outcome IRT model, known as a graded-response model (e.g., Treier and Jackman 2008, Pemstein et al 2010). This aggregation technique has the virtue of generating scores with meaningful interpretations. For example, a low score on a civil liberties scale might mean that a country provided freedom from torture and servitude but not any of the harder-to-achieve civil liberties; a relatively high score mean that it provided all of them except the most difficult, access to justice.

Latent variable measurement models will allow us to specify each indicator as either a cause or an effect of latent components, and the components as reflections of more general latent conceptions (Bollen 1993, Treier and Jackman 2008, Pemstein et al 2010). Specifically, we will use a new technique, Bayesian latent-variable modeling with Monte Carlo Markov Chain (MCMC) estimation, which was developed by Drew Linzer for a judicial independence application with Jeffrey Staton (Linzer & Lewis 2011; Linzer and Staton 2011). This technique is especially useful for constructing scales from time-series, cross-sectional indicators that are bounded at the extremes, have different numbers of ordinal thresholds, and are missing observations. A variant of this kind of measurement model could use the completely disaggregated, expert-level data in some of these models, which would enable us to model and correct for sources of expert bias. Our online questionnaire includes a short battery of questions about aspects of the experts' background that we suspect may affect their judgment: gender; age; education; type of employer; and country of citizenship, residence, education, and employment.

Latent-variable models can also be used to answer additional questions such as the following:

How reliable are the country experts, and why? How much weight should each indicator be given when measuring each component, and how much weight should each component have when measuring each conception? Are ratings more reliable for recent years than for earlier years? How much more reliable are ratings of well-known countries compared to obscure countries? Are simpler indices good approximations of the most rigorous ones? Which components of democracy are measured by extant indices, and how well? And perhaps most importantly, how far can the number of dimensions of democracy be reduced without sacrificing information about theoretically important differences, what is the nature of the dimensions that make both conceptual and empirical sense, and how have these dimensions evolved over the past century?

Clarifications

Several important clarifications apply to the V-Dem conceptual scheme. First, although all components and indicators must be closely associated with at least one of the seven conceptions of democracy (see Table 1), they may be associated with more than one. Accordingly, a given indicator may be relevant for more than one component. For example, a free and independent press enhances the competitiveness of elections (electoral democracy), the rule of the majority (majoritarian democracy), and the meaningful participation of citizens (participatory democracy). That said, a free and independent press is directly and definitionally entailed by the liberal model of democracy, which is why it is granted greater priority within that principle of democracy.

Second, it is important to acknowledge that conceptions and higher-order components, while much easier to define than *democracy*, are still resistant to authoritative conceptualization. Each might be defined minimally (a "thin" set of attributes) or maximally (a "thick" set of attributes).[27] A thin concept of electoral democracy, for example, might include only elective government, free and regular elections. A thicker concept of electoral democracy might add male and female suffrage, party strength, party system stability, electoral system proportionality, turnover, competitiveness, subnational government elections, and even civil liberty and media development. Our objective in Appendix B is to identify the attributes we imagine most people would associate with minimal and maximal definitions. Even so, we are keenly aware that others might make different choices, and that different choices will be required for different tasks. The purpose of the taxonomy is guidance, not legislation. It demonstrates how the various elements of V-Dem hang together, according to a particular set of inter-relationships. We expect other writers will assemble and dis-assemble these parts in whatever fashion suits their needs and objectives. In this respect, V-Dem has the quality of a Lego set.

Third, some components and indicators are undoubtedly more important in guaranteeing a polity's overall level of democracy than others. This of course depends upon one's model of democracy. In any case, inclusion in V-Dem does not presuppose a judgment of relative importance. All it means is that a particular component or indicator is relevant to the operationalization of at least one conception of democracy.

Fourth, principles, components, and indicators of democracy sometimes conflict with one another. At the level of principles, there is an obvious conflict between the liberal and majoritarian principles, which adopt contrary perspectives on most institutional components: fragmented power satisfies the liberal ideal but inhibits the possibility of majority rule. One can easily perceive conflicts across other conceptions as well.

Similar conflicts are in evidence at lower levels of aggregation. For example, protection of individual liberties can impose limits on the will of the majority; and the existence of strong civil society

[27] See Gerring (2012: ch 5) on minimal and maximal definitions and Coppedge (1999) on thick and thin definitions.

organizations can have the effect of pressuring government to restrict the civil liberties enjoyed by marginal groups (Isaac *n.d.*).

Furthermore, the same institution may be differently viewed according to different conceptions of democracy. For example, the common practice of mandatory voting is clearly offensive to the liberal model (where individual rights are sacrosanct and include the right not to vote), but is vindicated by the participatory model (since it has a demonstrated effect in boosting turnout where sanctions are more than nominal).

Such contradictions are implicit in democracy's multidimensional character. No wide-ranging empirical investigation can avoid conflicts among democracy's diverse attributes. However, with separate indicators representing these different facets of democracy it will be possible to examine potential tradeoffs empirically—an important substantive issue for policymakers and academics to grapple with.

Finally, our proposed set of components and indicators, while fairly comprehensive, is by no means entirely comprehensive. The protean nature of *democracy* resists closure; there are always potentially new components/indicators that, from one perspective or another, may be associated with this essentially contested term. Moreover, some conceptions of the concept are difficult to capture empirically, and virtually impossible to track over time and across countries on a global scale. This limits the scope of any empirical endeavor.

III. Data Collection

The viability of any dataset hinges critically on its method of data collection. V-Dem aims to achieve transparency, precision, and realistic estimates of uncertainty with respect to each data point. In these respects, as in others, we believe that V-Dem compares favorably with extant democracy indicators. In order to assess these issues, however, it is important to lay out our approach in some detail.

The nearly 400 indicators listed in Appendix A fall into three types. Type (a) data is gathered from extant sources (other datasets or secondary sources) and requires no original coding. The collection of this data is supervised by the project managers. Sources are listed in Appendix A.

Type (b) data is gathered from country-specific sources and does not require coding decisions (being factual in nature). This data is gathered by country research assistants under the supervision of project managers and regional managers.

Type (c) data requires some degree of judgment about the state of affairs obtaining in a particular country. This genre of indicator is coded by country experts – generally academics or policymakers who are nationals of or residents in the country in question, with deep knowledge of a particular country and a specific substantive area.

Sometimes we draw on more than one of these coding sources for a particular indicator, or we ask country experts to validate coding derived from some other source. Thus, the data collection mechanisms outlined above – (a), (b), and (c) – may be combined, as listed in Appendix A. The table at right shows the number of indicators to be coded in each way.

type	number of indicators
a	103
b	59
c	154
a&b	6
b&c	25
a&b&c	52
total	399

Country Expert Coding

Type (c) coding – by country experts – is the most difficult but also the most critical to the project. A vast majority of the data falls into this category – a total of 231 indicators. This represents a unique set of

original data over 112 years. Accordingly, a number of steps are taken to minimize error and to gauge the degree of imprecision that remains.

Most important, we endeavor to craft questions in such a manner that their meaning is clear and specific and not open to a wide variety of interpretations. They should mean the same thing (more or less) in each context and not suffer from temporal or spatial non-equivalence. The V-Dem pilot test (described below) served as an initial test of our questionnaire, and prompted quite a few revisions in the next round of surveys.

Five experts are enlisted to code every indicator for each country. The process works as follows. First, we identify a list of potential coders for a country, drawing on those known to the regional manager or to members of the project team, or on recommendations by those we know. For each person, we compile basic information – country of origin, current location, highest educational degree, current position, area of expertise, and any possible biases that might affect their ability to code questions in a dispassionate manner. This list is then culled for most likely prospects – those who are judged to have the requisite knowledge, equanimity, and diligence to code each area of the survey. All those who are eventually contacted are asked for additional names (a "snowball" sampling technique). In this fashion, we are able to generate a large number of potential coders.

Where we fear that experts might be divided on how to code a country, and where the division seems reasonable – i.e., reasonable people may disagree – we endeavor to choose coders who represent different views and political traditions of the country in question. For example, experts chosen for Venezuela will include a mix of experts from both the pro- and anti-Chávez camps.

A number of steps are taken to assure informed consent and confidentiality among participants in the project. The on-line survey provides full information about the project (including this document) and the use of the data, so that coders are fully informed. It also requires prospective coders to choose one of three levels of confidentiality and certify that they accept the terms of the agreement. They can access the surveys only with a username and password that we assign. Their data is stored on a firewall-protected server managed by the University of Notre Dame. Any data released to the public that includes information about experts will exclude information that might be used to identify coders.

Each country expert is responsible for one or several sections of the survey – those with which s/he is most familiar. Categories include sovereignty, voting and representation, elections, the executive, the legislature, the judiciary, political parties and party system, the media, civil society associations, direct democracy, subnational politics, civil liberty, equality, and deliberation, as listed in Appendix A. These subdivisions are especially important for highly specialized realms such as the judiciary, whose questions are difficult for non-specialists to code. This means that each country is analyzed by anywhere from 5 to 15 country experts.[28]

Contact between V-Dem and country experts is primarily through email (though Skype calls are sometimes helpful). Initial invitations are issued by the regional manager in charge of that country. Subsequent communication may be from the regional manager or the central administration (headquartered in at the universities of Gothenburg and Notre Dame). The most common sort of communication concerns (a) answering questions about the process (how to access the web site and work through the on-line survey), (b) nudging experts to complete the survey, and (c) arranging for compensation. Coders receive a modest honorarium for their work, in accordance with the number of questions they have answered.

Once a survey (i.e., a section of the questionnaire) has been completed, results are reviewed by

[28] Note that in some small and under-studied countries (e.g., Yemen) it has been necessary to ask individual experts to code the whole set of surveys, simply because experts on the various parts of the survey are not available. But this is rare.

the regional manager and by research assistants at project headquarters. In rare circumstances, where a particular coder's responses are grossly out of line with other coders' responses to a question, or where the responses show evidence of fatigue or inattention (e.g., where the same coding is repeated for a century, even in the face of obvious changes in the country's history), that coder will be contacted and asked to reconsider. A short conversation about the meaning of the question may ensue, at which point the coder will have the opportunity to revise his/her codings. However, for the most part, decisions by coders are respected. We fully expect that for difficult questions – those for which a factual basis for coding is not available – a range of responses is possible. It is desirable, under the circumstances, for the coding results to reflect this uncertainty.

Experts are required to report a level of confidence for each coding, an indication of their subjective level of uncertainty. This is scored on a scale from 0 to 100.

A second opportunity for registering uncertainty is in the "Remarks" fields that lie at the end of each section. Here, experts can comment (in prose) on any aspect of the indicators or codings that s/he found problematic or difficult to interpret.

Diagnostics

After coding is complete, a number of diagnostic tests will be conducted in order to gauge whether there might be some form of systematic bias in the coding. Simple tests can be done by regressing a responses for each question against various characteristics of the coder along with country and year dummies, then observing whether any of these characteristics are significant predictors. A similar test can be done at higher levels of aggregation by including the coder characteristics as predictors in more fully specified mulitlevel measurement models for components. This procedure can be repeated across all questions of type (c). Relevant background features include: (1) sex, (2) age, (3) highest degree received, (4) location of graduate work (Anglo-Europe or not), (5) place of birth (in-country or abroad), (6) years spent in country, (7) years spent away from country, (8) location of employer (in country or not), (9) employer (a 9-part classification, tested as a series of dummies), (10) overall estimate of level of democracy in country that has been coded, (11) overall estimate of level of democracy in selected countries around the world (understood as separate predictors), (12) stance towards free markets, (13) preferred models of democracy, (14) stance toward the current regime in the country that has been coded.

Several of these questions deserve further explication. #10 derives from a question that asks coders to assign a point score on a 0-100 scale summarizing the level of democracy in the country they are coding about which country in the world the coder perceives as most democratic, overall (using whatever understanding of democracy they themselves prefer). This will be transformed into standard deviations from the mean (Z scores) for a given country and then combined into a common variable – differentiating which coders feel more optimistic about democracy in their country from those who feel more pessimistic.

#11 is similar, except that here we select prominent countries from around the world that embody varying characteristics of democracy/autocracy and ask coders to provide an estimate of the overall level of democracy in these countries (on the same 100-point scale). Chosen exemplars include Costa Rica, Cuba, India, Nigeria, North Korea, Russia, Saudi Arabia, South Africa, Sweden, Switzerland, the United Kingdom, the United States, and Venezuela. Answers to each question constitute a separate interval variable, generating twelve variables (which should not be tested at the same time due to collinearity).

#12 gauges the coder's views of the relationship between markets and democracy. "It is sometimes argued that free markets and strong protection for property rights are essential for the

establishment of democracy." The coder is then asked whether s/he agrees or disagrees with this statement, along a Likert scale.

#13 asks explicitly about the coder's views of democracy. We list and briefly explain the six conceptions of democracy outlined in a previous section of this document – electoral, liberal, majoritarian, participatory, deliberative, egalitarian – asking coders to indicate their sympathy for them on a 5-point Likert scale.

#14 attempts to gauge the coder's relationship to the current regime in the country they are coding. We do not ask this question of the coder but rather of the regional manager, who is expected to have some familiarity with the country expert s/he chooses to work on V-Dem. Once the coder has accepted responsibility for coding, but prior to beginning coding, the RM will assess the status of the coder, as best as can be determined. Because this is a sensitive question, we insert the full question here. "How would you characterize the country expert's relationship to the party or individual that currently controls the most powerful office in the country? In some instances, this may be readily apparent from publicly expressed views. In other instances, it may be less apparent but it is possible to guess – from the person's position, family background, ethnic group, and so forth. We regard this as a best guess rather than a factually grounded account. Its only purpose is to test for possible problems of systematic bias in the coding. Answers will not be revealed to the coder; nor will they be accessible to the general public." Possible answers take the form of a Likert scale.

For further details on these questions see Appendix A.

Combining Expert Ratings

Having five coders for each type (c) question is immensely useful, as it allows us to identify wayward coders as well as to conduct inter-coder reliability tests. (As noted, these sorts of tests – standard in most social science studies – are rarely if ever employed in extant democracy indices.) However, we also need to combine these ratings in order to provide a single "best estimate" for each question.

To combine expert ratings for a particular country/indicator we employ a well understood and robust technique, the linear opinion pool (O'Hagan et al. 2006, 181-184). This procedure gives greater weight to scores about which the experts feel more confident. For interval-level data, this would be a weighted average based on the experts' prior beliefs that their ratings are correct. Most of our indicators, however, are ordinal or nominal, so combinations are a bit more complicated. We calculate the priors for each scale level surrounding an expert's rating, based on the confidence level the expert assigned to his or her preferred rating. We then calculate a weighted sum of these priors for all the experts who answered the question for that country-date. The combined rating is the one with the highest total probability. Naturally, we also report what that probability is – which may be useful for subsequent analyses.

The priors are calculated from the experts' confidence levels and the number of levels in each variable. For a two-level scale, zero confidence corresponds to 50-50 probabilities; and to a similarly flat distribution for scales with multiple levels. For nominal items, the priors for non-selected scores are distributed evenly among all the alternatives; for ordinal items, they favor the nearest alternatives over those that are farther from the preferred score. Table 2 illustrates the procedure. Although much of the math is hidden in this table, one can appreciate that preferred scores have the highest prior for each expert, adjacent scores have priors that diminish with distance, high-confidence ratings have greater weight, and the summary score (2) is the one with the greatest sum of weighted priors, which sum to 1.0. The two shaded cells contain the information that would be reported for this observation in the dataset. We will also analyze inter-coder reliability using more familiar statistics such as Cronbach's alpha and other versions of the intra-class correlation coefficient (ICC).

Our data will be the first democracy data (and among the first in national-level comparative social science) to include score-specific reliability estimates. Reliability estimates can be used in several ways. They will enable users to select only the most reliable data. They can be used to weight analyses to correct for heteroscedasticity when our indicators are the dependent variables and reliability is correlated with a regressor. They will also guide the project team as we improve the data over time, indicating which survey items need modification or coding by additional experts. And they can be used to improve index construction, as discussed below.

IV. Project Management

What follows is a brief overview of the main components of V-Dem – the project team, the advisory board, the web site, sustainability into the future, and outreach to the international community.

Project team

A complex undertaking such as this requires coordinating the work of many individuals, each of which brings a unique skill or knowledge base to the project. *Principal investigators* – Coppedge, Gerring, Lindberg, and Teorell – manage the project team and oversee all activities.

Project managers assume responsibility for at least one category of indicators, as follows: Sovereignty (Bernhard), Voting and representation (Paxton), Elections (Lindberg), The executive (Teorell), The legislature (Fish/Kroenig), The judiciary (Staton), Political parties and the party system (Hicken), The media (Semetko), Civil society associations (Bernhard), Direct democracy (Altman), Subnational politics (McMann), Civil liberty (Skaaning), Equality (Gerring), and Deliberation (Gerring). This division of labor enabled us to define concepts, identify existing data, and write survey questions that reflect the best thinking on each theme, in consultation with other experts in these areas. Project managers are also responsible for assuring cross-country equivalence for their indicators, so that a given question is not interpreted differently in varying country contexts.

Linzer serves as a V-Dem's *measurement expert* and will be responsible for developing reliability tests and aggregation techniques, as discussed in previous sections.

Regional managers are responsible for recruiting and managing expert coders, overseeing coding, and checking for problems of equivalence at the regional level. Regions include: North America/Caribbean, Central America, South America (Carlos Gervasoni, Department of Political Science and International Studies, Torcuato DiTella University, Argentina), Francophone West & Central Africa, Anglophone East & Central Africa (Samuel Afolabi Olugbemiga, Obafemi Awolowo University, Ile Ife, Nigeria), Lusophone Africa, Southern Africa, North Africa, Northwestern Europe, Southern Europe (Tiago Fernandes, Department of Political Studies, New Lisbon University, Lisbon, Portugal), Central Europe, Russia/Central Asia, Middle East, South Asia, East Asia, Southeast Asia, and Pacifica. Personnel are indicated in parentheses, in cases where this position has been filled.

Expert coders are country experts who code type (c) indicators. They are generally holders of a PhD and usually residents or citizens of the country they are coding. They also have specialized knowledge in at least one of the indicator categories (Sovereignty, Voting and representation,…).

Research assistants will assist at all levels except expert coding. They are generally pursuing graduate studies, and will work under the supervision of members of the project team. A *Country research assistant* will be hired for each country, to gather the type (b) data for that country. These RAs

work under the supervision of the Regional Managers but can also work with Principal investigators and Project managers.

Advisory board

V-Dem's advisory board includes a mix of academics, jurists, and practitioners from around the world. Board members provide guidance on all aspects of the project. While communication is normally conducted through email and Skype, the board will meet periodically in the coming years to facilitate in-depth discussion. (The advisory board is not responsible for the content or conduct of V-Dem.)

Aitzaz AHSAN is Senior Advocate of the Supreme Court of Pakistan. In 2007-08 he was president of Pakistan's Supreme Court Bar Association. During 1988-93 he was a member of the National Assembly. In 1994-99 he was a member of the Senate, where he was Leader of the Opposition. He has also served as Minister of the Interior and as Minister of Law and Justice.

Abdullah M. ALSAIDI is Senior Fellow at the International Peace Institute. From 2002 until 2011 he was Permanent Representative of Yemen to the United Nations. He served as Yemen's Vice-Minister of Foreign Affairs from 1999 to 2002, and as a member of the National Arbitration Committee from 1996 to 2002.

Larry DIAMOND is Senior Fellow at the Hoover Institution. He is also a senior fellow at the Freeman Spogli Institute for International Studies, Stanford University, where he is the director of the Center for Democracy, Development, and the Rule of Law (CDDRL). He is founding coeditor of the *Journal of Democracy*.

Sonia FLEURY is Professor of Political Science at the Brazilian School of Public Administration and Business, Rio de Janeiro. She has served as a member of Brazil's National Council on Economic and Social Development and on the Brazilian Health Social Determinants Commission.

Emmanuel GYIMAH-BOADI is Professor of Political Science at the University of Ghana and Executive Director of the Ghana Center for Democratic Development (CDD-Ghana). He is also codirector of the of the Afrobarometer project.

Cyd HO Sau-lan is a member of the Legislative Council of Hong Kong. She is also chair of Hong Kong Human Rights Monitor, a founding councilor of the World Future Council, and co-convener of Project Civil Referendum.

Niraja Gopal JAYAL is Professor, Centre for the Study of Law and Governance, Jawaharlal Nehru University, New Delhi. She has also served as Chair of the Centre and as Director of the Jawaharlal Nehru Institute of Advanced Study. She has held visiting fellowships at Princeton University, University of Melbourne, and the Maison des science de l'homme.

Jeremy K. B. KINSMAN is Director of the Community of Democracies project, Regent's Lecturer at the University of California, Berkeley, and Diplomat-in-Residence at the Woodrow Wilson School at Princeton University. He served as Canadian High Commissioner to the United Kingdom in 2000-02 and Canadian Ambassador to the European Union in 2002-06.

Augustin LOADA is Professor of Law and Political Science at the University of Ouagadougou, Burkina Faso. He is also Executive Director of the Centre pour la Gouvernance Démocratique (CGD), Ouagadougou.

Soledad LOAEZA is Professor of Political Science at the Center for International Studies, El Colegio de México, Mexico City. She is also a member of the Board of Advisers of the International Institute for Democracy and Electoral Assistance (IDEA).

Pratap Bhanu MEHTA is president of the Centre for Policy Research, New Delhi. He is a member

of the Government of India's National Security Advisory Board and has served as a member-convener of the prime minister of India's National Knowledge Commission.

Chung-in MOON is Professor of Political Science, Yonsei University, Seoul, and Editor-in-Chief of *Global Asia*. He has also served as Ambassador for International Security Affairs at the Ministry of Foreign Affairs and Trade and Chairman of the Presidential Committee on Northeast Asian Cooperation Initiative, Republic of Korea.

Saiful MUJANI is Executive Director and cofounder of Lembaga Survei Indonesia (LSI, Indonesian Survey Institute) and Professor of Social and Political Sciences, State Islamic University of Jakarta. He is also Director of Research at the Freedom Institute, Jakarta.

Ziya ÖNİŞ is Professor of International Political Economy and Director of the Center for Research on Globalization and Democratic Governance (GLODEM), Koç University, Istanbul. He has also served as Director of the Graduate School of Social Sciences and Humanities at Koç University and Professor of Economics at Boğaziçi University.

OYUN Sanjaasuren is a member of the Great State Assembly (parliament) of Mongolia, Chair of the Zorig Foundation, and President of the Mongolian Geological Association. In 2004-05 she served as the Vice-Speaker of the Great State Assembly. In 2007-08 she was Mongolia's Minister of Foreign Affairs.

Lilia SHEVTSOVA is Senior Associate and Chair of the Russian Domestic Politics and Political Institutions Program at the Carnegie Endowment for International Peace, Moscow Center. She is also Professor of Political Science, Moscow State Institute of International Affairs of the Ministry of Foreign Affairs of the Russian Federation.

G. J. W. VAN OVEN is Vice-President of the Court of Appeal in the Hague and President of the Judges for Judges Foundation. Between 1994 and 2003 he was a member of the States-General (parliament) of the Netherlands. From 1983 to 1990 he served as a public prosecutor in Amsterdam.

Web site

V-Dem is stored in an interactive web site (www.v-dem.net). The survey interface and underlying relational database were designed, following our guidance, by the Center for Research Computing at the University of Notre Dame. All data is entered on-line (exceptions are made for those country experts who might not have internet access) and will ultimately be available on-line.

All data (except confidential information about the coders) will be made public when coding for all countries is complete. (Data for selected countries may be released prior to the completion of the project.)

The site includes web-based interfaces that will soon allow users to download pilot data and to visualize these data in line graphs or histograms. It is being expanded to allow visitors to select variables, countries, and dates for the data download and to visualize the pilot data in scatterplots and maps as well. When the data collection is finished, all of the data will be publicly available through these interfaces.

After coding is complete, we plan to establish an open and transparent system of display and dissemination. This might include a blog or Wiki-style format in which interested individuals could comment on the scores provisionally assigned to a country that they know well. This commentary might take the form of additional information—perhaps unknown to the country expert—that speaks to the viability of the coding. Or it might take the form of extended discussions about how a particular question applies to the circumstances of that country. Naturally, a few cranky participants may be anticipated. However, the Wikipedia experience suggests that there are many civic-minded individuals, some of

them quite sophisticated, who may be enlisted and may have a lot to add to the discussion. At the very least, it may provide further information upon which to base estimates of uncertainty. At best, it will help to refine the V-Dem questionnaire in future iterations of the project.

Sustainability

V-Dem must be continually revised and updated in order to remain useful for scholars, practitioners, and citizens. Users will be encouraged to contribute suggestions for revision and these suggestions will be systematically reviewed. The set of indicators comprising V-Dem will doubtless be expanded, and some questions may require revision. If the latter is required, revisions will need to be implemented systematically (for all countries and years in the database) so that equivalence across countries and through time is maintained. Most important, V-Dem will need to be iterated on an annual or semiannual basis so that the database is reasonably up to date. We expect that it would be significantly less work for coders to update scores for the most recent year (or few years--it would be good to have them replicate some of the coding previously done, to check reliability). We would employ the same coders whenever possible for these updates, although some turnover is unavoidable.

International Collaboration and Outreach

The current project is international in character and composition to a much greater extent than any other democracy index. V-Dem has dual headquarters on either side of the Atlantic – at the Kellogg Institute at the University of Notre Dame, in Indiana, and at the V-Dem Institute at the University of Gothenburg, Sweden. The project team includes nationals of Denmark, Sweden, the United States, and Uruguay. The advisory board includes nationals of Brazil, Burkina Faso, Canada, Ghana, Hong Kong, India, Indonesia, Mexico, Mongolia, the Netherlands, Pakistan, Russia, South Korea, Turkey, the United States, and Yemen. Regional advisors will consist of eighteen academics situated in regions around the world, as listed above. Research assistants hired to work on various aspects of the project will be drawn from multiple countries. Altogether, members of the project team and the advisory board connect the V-Dem project to several dozen universities located in every region of the world. Most important, V-Dem enlists experts from every country in the world to serve as coders for our questionnaire.

We anticipate that the direct involvement of academics and professionals in V-Dem will result in numerous projects, many of which are collaborative in nature. Of course, the most important collaborations will arise after the project is complete. We anticipate that the V-Dem database and web site will become a focal point for policymakers, practitioners, and scholars across the social sciences.

In order to facilitate this we plan to introduce the project at social science conferences throughout the world. This includes the International Political Science Association meetings and meetings of regional political science associations. Project team members and members of the advisory board who live or work in each region will assist in that region's conferences.

Coincident with these academic events, we will ask our sponsors in the Swedish Ministry of Foreign Affairs to help host events to which democracy activists, professionals, and government officials are invited. Publications in high-profile academic journals and press outlets of a more popular nature will ensure that all relevant communities of interest are informed of the availability of V-Dem and how it can be used to help them do their work better. We hope to establish a working paper series at the Varieties of Democracy Institute at the University of Gothenburg and to encourage all users of the data to contribute their initial publications to it.

Traffic to the web site, as channeled by Google and other search engines and links on related

sites, offers a further channel for outreach – making V-Dem accessible to anyone who searches on key terms (e.g., "democracy"). Funding permitting, we plan to commission the creation of a V-Dem App for smart devices that would enable users to check scores for selected countries, variables, and years, and even to generate color-coded graphs and maps on demand.

V. Workplan and Funding

The progress of V-Dem may be summarized in three stages (which are not strictly sequential, as there is some back and forth between the first two stages).

First, there is the conception and management of the project itself. This stage is substantially complete, though we are still working to finalize the set of indicators listed in Appendix A.

Second, there is the data collection stage. We are now nearly finished with a pilot study of twelve countries and about to begin a second stage of data collection focused on twenty or so additional countries (dubbed the "Oslo" phase), as described below. Subsequent phases will focus on specific regions and need not occur simultaneously. (Indeed, we have found it desirable to focus on a handful of countries at a time.) We anticipate the completion of coding across all countries in the next few years. This of course is contingent upon obtaining additional funding.

A third and final stage involves cleaning the data, preparing it for public access (through the V-Dem web site), reaching out to potential end-users (including the community of democracy practitioners), analyzing the data for descriptive and causal relationships, and writing reports for eventual publication (in scholarly journals and in more popular venues).

Pilot study

With the support of Sweden's Ministry of Foreign Affairs, the Riksbankens Jubileumsfond (Sweden), the Quality of Government Institute at the University of Gothenburg, and the Kellogg Institute and other bodies at the University of Notre Dame, we have been carrying out a pilot study in twelve countries around the world. These include one "easy" country and one "hard" country in each of six regions: Mexico and Suriname in the Americas, Japan and Burma in Asia, Russia and Albania in the post-Communist region, South Africa and Ghana in Africa, Egypt and Yemen in the Middle East, and Sweden and Switzerland in Europe.

The pilot study began May 30, 2011 and has progressed rapidly. Our team has recruited 90 experts, who have submitted 460,000 country-date-question-expert observation ratings to the database in just seven months, including some data on all 12 countries and some on all survey questions. The speed of data collection could be multiplied with relatively little effort by working on more countries at the same time.

Preliminary analyses indicate that the data already collected are of good quality. For example, we ask all experts not only to choose a response to each question but also to estimate their confidence in all their responses. The average summary certainty level in the pilot data is 84 percent, which means that based on the the agreement among experts and their confidence in their ratings, we can be, on average, 84 percent confident that the summary scores are correct. Cronbach's alpha averages .709 (conventionally considered in the "acceptable" range) for the 131 pilot-study type (c) questions for which we have enough data to calculate it. We are doing more analysis of the pilot study data to pinpoint the weakest questions and modify or drop them, improving reliability further before the second pilot phase begins. In addition, very preliminary analyses have shown that aggregated scores

discriminate well not only across countries and over time, but also across components. For example, a country can do well with respect to civil society but less well on administration of elections or legislative strength. Such distinctions are among the most important contributions of our approach.

The "Second" Pilot Phase

The next phase of V-Dem begins in January, 2012, and extends to 30 September, 2012. This phase is financed by Research Council of Norway through a collaborative arrangement with the PI, Håvard Hegre (Oslo University). This phase will include Kenya, Tanzania, Uganda, Ethiopia, Nigeria, Colombia, Venezuela, Brazil, Argentina, Taiwan, South Korea, the Philippines, Thailand, Lebanon, India, Afghanistan, and Kosovo and perhaps other countries to be determined.

Funding

Initial funding has been provided by the Kellogg Institute at the University of Notre Dame, the Quality of Government Institute at the University of Gothenburg, Sweden's Ministry for Foreign Affairs, the Riksbankens Jubileumsfond (Sweden), and the Research Council of Norway. Further funding from these sources may be forthcoming.

In addition, discussions have begun with other potential funders in Europe and North America including the European Commission, CIDA, DfID, Denmark's Ministry for Foreign Affairs, the European Endowment for Democracy (EED), Norway's Ministry for Foreign Affairs, and USAID. An application with the US National Science Foundation is pending.

VI. Validation

It would be irresponsible to propose a new set of indicators for democracy without some tests of their validity. A project of this magnitude therefore requires validation. To be sure, an informal process of validation will undoubtedly ensue once the project is complete. End-users will write reviews in academic journals and – we hope – will also contribute to the wiki blogs that we plan to develop for the V-Dem web site. Here, we want to provide the outlines of a more formal test, one that could easily be applied prior to the project's completion (given sufficient funding).

This exercise is focused at the meso-level (components), rather than the macro-level (principles and conceptions) or micro-levels (indicators). V-Dem principles and conceptions are so highly aggregated that any judgment of their validity rests heavily on the aggregation technique employed rather than the validity of the underlying data. Thus, while aggregation issues (discussed in Part VI) are critically important, they are less subject to empirical validation since they hinge on choices of conceptualization and aggregation that are largely pre-empirical in character.

V-Dem indicators, by contrast, are already subject to an internal process of verification by virtue of the fact that they are coded independently by multiple country experts. Inter-coder reliability tests, along with self-reported uncertainty, may therefore be viewed as partial validation of the results – allowing us to arrive at a measure of certainty for each point estimate. Naturally, one could also replicate the coding of indicators with a new set of coders. It is unlikely that this procedure would yield substantially different outcomes, but we plan to do such replications during future updates.

Consequently, the appropriate focus for validity testing is the component. (Note that components serve as the conceptual building blocks for all V-Dem conceptions; so in testing components we are also offering a meso-level test of the conceptions.)

One approach to validity testing compares V-Dem components with other indices that purport to measure similar concepts. Thus, the correlation between a preliminary V-Dem index of civil liberties and the Freedom House index of civil liberties is -.87 (a negative sign because in Freedom House data, higher scores mean "less free").

It is reassuring to discover that V-Dem components are not uncorrelated with other like-minded measures. However, techniques of convergent validity are limited in their utility for two reasons. First, we have serious doubts about the validity of many standard indices, for reasons discussed in Part I. Second, standard indices tend to hover at a higher level of aggregation, thus impairing comparability between V-Dem and non-V-Dem indices. Indeed, only a few extant indices are close enough in conception and construction to provide an opportunity for direct corroboration. Consequently, we must look elsewhere for validation of the V-Dem project.

We therefore plan to commission a validation study by a neutral body unconnected with V-Dem, such as an agency, an NGO, or an academic committee. This body would randomly select countries and ask highly qualified independent experts to examine V-Dem component scores for their country of expertise. The experts would be asked whether each component (and its certainty score) plausibly tracks changes over time within the country; whether the contemporaneous differences across components in the same country reflect its known political evolution; and whether cross-national differences also make sense. The neutral body would compile these expert evaluations and write a report that we would make available to the public.

VII. Payoffs

Having described V-Dem in some detail, the payoffs anticipated from our project may now be briefly enumerated.

We believe that this vast database of practically all aspects of democracy, for all polities, for more than a century of history, has the potential to help forge consensus on how to think about democracy. The term "democracy" itself is highly contested; having to agree on a single measure of democracy raises the stakes in what are basically semantic disagreements. By recognizing multiple democratic principles and conceptions and shifting the focus to lower-level components, we lower the stakes, making it easier to find concepts on which everyone can agree.

The importance of creating consensus on these matters can hardly be overemphasized. The purpose of a set of democracy indicators is not simply to guide rich-world policymaking bodies such as USAID, the World Bank, and the UNDP. As soon as a set of indicators becomes established and begins to influence international policymakers, it also becomes fodder for dispute in other countries around the world. A useful set of indicators is one that claims the widest legitimacy. A poor set of indicators is one that is perceived as a tool of western influence or a mask for the forces of globalization (as Freedom House is sometimes regarded). The hope is that by reducing the elements of democracy down to levels that are more coherent and operational it may be possible to generate a broader consensus around this vexed subject. Moreover, insofar as indicators might be used to construct aggregate indices, at least everyone would know precisely why they received a particular score in a particular year, since the underlying indicators used to compose the index would be available for inspection. (Presumably, the aggregation principles would be available as well.)

Another advantage is the degree of precision and differentiation that a disaggregated set of

indicators offers relative to extant composite indices. While holistic measures of democracy float hazily over the surface of politics, the conceptions, components, and indicators of a disaggregated dataset are comparatively specific and precise. Contrasts and comparisons become correspondingly acute.

Note that the data we are collecting is extremely disaggregated, down to the level of country-date-variable-expert observations. When completed, it will become one of the largest social science databases in existence. For anyone interested in issues relating to democracy, the data in this disaggregated form will be a vast resource for exploration and testing: a database of approximately 400 variables for 193 countries over 112 years, with about 5 experts per item for the 231 non-factual indicators (or approximately 1000 partially overlapping raters), resulting in about 28 million data points. V-Dem will therefore allow policymakers and researchers to clarify how, specifically, one country's democratic features differ from others in the region, or across regions.

This detailed knowledge of cases is especially helpful in the context of country assessments. How can policymakers determine which aspects of a polity are most in need of assistance? While Freedom House and Polity offer only a few dimensions of analysis (and these are highly correlated and difficult to distinguish conceptually), V-Dem envisions numerous parameters. It seems clear that for assessing the potential impact of programs focused on different elements of a polity it is helpful to have indicators at hand that offer a differentiated view of the subject. Intuitively, the greatest effectiveness is achieved when program interventions are targeted on the weakest element of democracy in a country. A large set of fully differentiated indicators would make it possible to both identify those elements and test the assumption behind such choices.

Relatedly, V-Dem will allow policymakers and researchers to track a single country's progress or regress through time. One will be able to specify which facets of a polity have improved, and which have remained stagnant or declined. This means that the longstanding question of regime transitions would be amenable to empirical tests. When a country transitions from autocracy to democracy (or vice versa), which elements come first? Are there common patterns, a finite set of sequences, prerequisites? Or, is every transition unique? Do transition patterns effect the consolidation of democracy? With a large set of indicators measured over many years, it would become possible for the first time to explore transition sequences.[29] Does a newly vibrant civil society lead to more competitive elections, or to an authoritarian backlash? Do accountable elected officials create an independent judiciary, or does an independent judiciary make officials accountable? Similar questions could be asked about the relationships among citizenship, voting, parties, civil society, and other components of democracy, perhaps with the assistance of sequence-based econometrics (Abbott 1995; Abbott and Tsay 2000; Wu 2000).

Note that insofar as one wishes to judge trends, trend lines are necessary. A single snapshot of the contemporary world reveals nothing about the direction or speed at which countries are moving toward, or away from, democracy. Even trends in a short span of recent years can be very misleading, as many democratization paths contain many years of stasis punctuated by sudden movements toward or away from democracy. Assessments of global trends require even more data, as some countries move in opposite directions in any given year; "waves" of democratization exist only on average, with many exceptions (Huntington 1991).

Policymakers also wish to know what affect their policy interventions might have on a given country's quality of democracy. There is little hope of answering this question in a more than suggestive fashion if democracy is understood only at a highly aggregated level or over a small number of years (Finkel, Pérez-Liñán, and Seligson 2007). The intervention is too small relative to the outcome to draw

[29] Sequencing is explored by Schneider and Schmitter 2004b with a smaller set of indicators and a shorter stretch of time. See also McFaul 2005.

strong causal inferences between USAID policies, on the one hand, and Country A's level of democracy (as measured, e.g., by Freedom House or Polity) on the other. However, it is more plausible to estimate the causal effects of a program focused on a particular element of democracy if that element can be measured separately and over a long period of time. Thus, election centered programs might be judged against an outcome that measures the quality of elections (Kelley 2009). This is plausible, and perhaps quite informative—although, to be sure, many factors other than international actors affect the quality of elections in a country. (There are many potential confounders.) The bottom line is this: insofar as policymakers must make reference to country-level outcome indicators, they will be much better served if these indicators are available at a disaggregated level. (Hopefully, they will also have even more fine-grained outcome measures with which to evaluate the efficacy of programs.)

One would also be able to test democracy's causal effect as an independent variable. Does democracy hinder economic growth, contain inflation, promote public order, or ensure international peace? Answering such questions requires a lengthy time-series because the effects of these factors play out over many years. They also require a great deal of disaggregation because we need to know, as specifically as possible, which elements of democracy are related to which results. This is helpful from the policy perspective as well as from the analytic perspective, so that we can gain insight into causal mechanisms. Whether democracy is looked upon as an independent (causal) variable, or a dependent (outcome) variable, we need to know which aspect of this complex construct is at play.

Recent work has raised the possibility that democracy's effects are long-term, rather than (or in addition to) short-term (Gerring et al. 2005). It seems quite probable that the short-term and long-term effects of democracy are quite different. Plausibly, long-term effects are more consistent, and more positive along various developmental outcomes, than short term effects. Consideration of these questions demands a historical coding of the key variables.

A final set of advantages of our proposed approach to democracy derives from its data-collection strategy. This is open to public scrutiny and commentary. Procedures of collection and aggregation are replicable. Experts with deep knowledge of each country's history are to be enlisted. Several experts are asked to code each indicator wherever the latter requires a degree of judgment. Inter-coder reliability tests, along with other gauges of reliability, are systematic. Periodic revisions are planned rather than ad hoc, and where re-coding is required it is applied to the whole dataset so that consistency is preserved through time and across polities.

With democracy indicators, as with most things, the devil is in the details. Getting these details right, and providing full transparency at each stage, should enhance the precision and validity of the measurement instrument, as well as its legitimacy in the eyes of potential end-users.

Conclusions

All of these factors should enhance the utility of the proposed approach to conceptualization and measurement. Moreover, we anticipate that the payoffs will be as vital for policymakers as for academics. Both of these constituencies are engaged in a common enterprise, an enterprise that has been impeded by the lack of a sufficiently discriminating instrument.

In this context, one might ponder the sort of problems that would arise for macroeconomists, finance ministers, and policymakers at the World Bank and IMF if they possessed only one highly aggregated indicator of economic performance: an index of "prosperity," for example. As good as GDP is (and there are of course difficulties), it neither measures the whole concept of prosperity nor says anything specific. It would not go very far without the existence of additional variables that measure the

components of this macro-level concept. This is in fact where the economics discipline stood 80 years ago: aside from census data, which generated some information about income and employment once a decade, and some trade figures (because trade was taxed), there were only "boutique" economic indicators compiled by lone academic economists for specific research projects. This situation changed with the Great Depression, when Simon Kuznets was appointed to set up an office in the Department of Commerce that would collect national accounts data in a systematic fashion (Carson 1975). In the field of political analysis, we are where economists were before the Depression. We have some crude sense of how democratic a country is; but we have no systematic knowledge of how a country scores on the various components of democracy, and our historical knowledge is even weaker.

Now let us step back from the details of our project to ask a larger question in the sociology of knowledge. If a new set of democracy indicators promises so many returns one might wonder why it has not already been developed. After all, academics and policy makers have been struggling with issues of conceptualization and measurement for quite some time, and the problems noted at the outset of this paper are widely acknowledged.

It is important to remember that producing a dataset of this immense scope is time-consuming and expensive, requiring the participation and coordination of many researchers. While the downstream benefits are great, no single scholar or group of scholars has the resources or the incentives to invest.[30] Indeed, the academic disciplines do not generally reward members who labor for years to develop new data resources.

Moreover, few national or international organizations have the funding and the motivation to collect global and historical data on a highly disaggregated level. While a host of national institutions (e.g., finance, commerce, agriculture, and labor departments, along with national banks) and international financial institutions (e.g., the World Bank, International Monetary Fund, World Trade Organization, Food and Agriculture Organization, and Organization for Economic Cooperation and Development) supervise the collection of economic data, there are no analogous institutions charged with collecting political data.

Thus, the topic of democracy has been left to several meagerly financed organizations who must go begging each year (or every several years) to renew their funding so that coding can be continued.[31] Freedom in the World, including the construction of the PR and CL indices (along with associated staff, consultants, printing, and promotion costs), is budgeted at $500–600,000 annually.[32] Polity consumes about $120,000 annually.[33] These are paltry sums, particularly when compared with the vast sums of money spent by OECD countries on democracy promotion.

A brief review of these expenditures may be helpful in putting our subject into the proper context. USAID spends about $600 million annually on its various democracy and governance programs.[34] Annual contributions to the United Nations' Democracy Fund add up to over $20 million.[35]

---

[30] Note that while scholars who are discontented with the leading indicators of democracy periodically re-code countries of special concern to them (e.g., Acuna-Alfaro 2005; Berg-Schlosser 2004a, 2004b; Bowman, Lehoucq, and Mahoney 2005; McHenry 2000), this re-coding is generally limited to a small set of countries or a small period of time. The incentive structure for data collection is discussed compellingly in Herrera and Kapur 2007.

[31] The following statistics are drawn from a variety of sources, as cited. They represent the most recent figures available. Where data for several recent years is available these are averaged together to form an estimate of normal budget activity.

[32] Arch Puddington (Director of Research, Freedom House), personal communication, January 2011.

[33] Monty Marshall, personal communication, December 2010.

[34] This figure, drawn from Finkel, Pérez-Liñán, and Seligson 2007, is for 2003 (in constant 1995 US dollars), and may have grown slightly since then.

[35] See www.un.org/democracyfund/Donors/donors_index.html.

DFID, the British development agency, spends $1.25 million annually on democracy and governance (though this also includes some anti-poverty programs).[36] DANIDA, the Danish development agency, spends $330 million annually in assistance to foster civil society and public administration.[37] SIDA, the Swedish development agency, spends $690 million annually to foster democratic governance and human rights.[38] CIDA, the Canadian development agency, spends $566 million annually for programs to improve the quality of government and civil society.[39] The European Commission, a body of the EU, spends $220 million annually for programs in support of government and civil society.[40] The Secretariat for Political Affairs (SPA) within the Organization of American States (OAS) spends $3.5 million annually to strengthen political processes in the member states, in particular to support democracy as the best option for ensuring peace, security, and development.[41]

Altogether, countries around the world target about $1.8 billion on projects related to government and civil society.[42] It is worth noting that rigorous evaluation of the impact of these programs would be enhanced by an investment of a small fraction of one percent of these expenditures, requisite for the creation of more disaggregated indicators of specific attributes of democracy, as outlined here.

In any case, the failure to adequately measure democracy is a product both of paltry resources and poor institutional incentives. Consequently, academics and policymakers have continued to employ—and complain about—Polity, Freedom House, DD, and other highly aggregated indices. It is our hope that users of these indices will recognize the public good aspect of enhanced measures of democracy, and that investments by individual scholars and funding institutions will help make this prospect a reality.

---

[36] This figure represents total bilateral gross public expenditure for DFID and other official UK sources. See www.dfid.gov.uk/About-DFID/Finance-andperformance/Aid-Statistics/Statistic-on-International-Development-2010/SID-2010—Tables-index/.

[37] See www.um.dk/en/menu/DevelopmentPolicy/DanishDevelopmentPolicy/AnnualReport/.

[38] SADEV (Swedish Agency for Development Evaluation): www.sadev.scb.se/eng/database/SADEV/bilateralt/bilateralt.asp.

[39] CIDA's Statistical Reports on International Assistance. See www.acdi-cida.gc.ca/acdi-cida/acdi-cida.nsf/eng/JUD-4128122-G4W#pre.

[40] European Commission Annual Reports on development policy and external assistance. See http://ec.europa.eu/europeaid/multimedia/publications/index_en.htm.

[41] The figure represents the mean value for 2005–2007. This line item disappears in 2008, absorbed within a broader category of spending. OAS Secretary General Annual Reports: www.oas.org/es/centro_informacion/informe_anual.asp.

[42] This figure represented committed funds (not funds actually spent) for 2009 (in constant 2000 USD). See Nielson, Powers, and Tierney 2010.

References

Abbott, Andrew. 1995. "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology* 21:93–113.

Abbott, Andrew, and Angela Tsay. 2000. "Sequence Analysis and Optimal Matching Methods in Sociology." *Sociological Methods and Research* 29: 3–33.

Acuna-Alfaro, Jairo. 2005. "Measuring Democracy in Latin America (1972–2002)." Working Paper No. 5, Committee on Concepts and Methods (C&M) of the International Political Science Association (August).

Altman, David, and Aníbal Pérez-Liñán. 2002. "Assessing the Quality of Democracy: Freedom, Competitiveness and Participation in Eighteen Latin American Countries." *Democratization* 9(2): 85–100.

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

American Political Science Association. 1950. *Toward a More Responsible Two-Party System*. New York: Holt, Rinehart.

Arat, Zehra F. 1991. *Democracy and Human Rights in Developing Countries*. Boulder: Lynne Rienner.

Arndt, Christiane, and Charles Oman. 2006. "Uses and Abuses of Governance Indicators." Paris: OECD Development Centre.

Arneson, Richard J. 2004. "Democracy Is Not Intrinsically Just." In *Justice and Democracy*, ed. Keith Dowding, Robert E. Goodin, and Carole Pateman. Cambridge: Cambridge University Press.

Bächtiger, André, ed. 2005. Special issue: "Empirical Approaches to Deliberative Democracy." *Acta Politica* 40(1–2).

Bagehot, Walter. 1963 [1867]. *The English Constitution*. Ithaca: Cornell University Press.

Barber, Benjamin. 1988. *The Conquest of Politics: Liberal Philosophy in Democratic Times*. Princeton: Princeton University Press.

Beetham, David, ed. 1994. *Defining and Measuring Democracy*. London: Sage.

_____. 1999. *Democracy and Human Rights*. Cambridge: Polity.

_____. 2004. "Towards a Universal Framework for Democracy Assessment." *Democratization* 11(2):1–17.

Beetham, David, Sarah Bracking, Iain Kearton, and Stuart Weir, eds. 2001. *International IDEA Handbook on Democracy Assessment*. The Hague: Kluge Academic Publishers.

Benelo, George C. and Dimitrios Roussopoulos, eds. 1971. *The Case for Participatory Democracy: Some Prospects for a Radical Society*. New York: Viking Press.

Berg-Schlosser, Dirk. 2004a. "Indicators of Democracy and Good Governance as Measures of the Quality of Democracy in Africa: A Critical Appraisal." *Acta Politica* 39(3): 248–278.

_____. 2004b. "The Quality of Democracies in Europe as Measured by Current Indicators of Democratization and Good Governance." *Journal of Communist Studies and Transition Politics* 20(1): 28–55.

Berman, Sheri. 1997. "Civil Society and the Collapse of the Weimar Republic." *World Politics* 49(3): 401–29.

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

Bertelsmann Foundation. [various years]. *Bertelsmann Transformation Index [2003–]: Towards Democracy and a Market Economy*. Washington, DC: Brookings Institution Press.

Besancon, Marie. 2003. "Good Governance Rankings: The Art of Measurement." World Peace Foundation Reports 36. Cambridge, MA: World Peace Foundation.

Bohman, James. 1998. "Survey Article: The Coming of Age of Deliberative Democracy." *Journal of Political Philosophy* 6(4): 400–425.

Boix, Carles and Sebastian Rosato. 2001. "A Complete Data Set of Political Regimes, 1800–1999." Chicago: Department of Political Science, University of Chicago.

Boix, Carles, and Susan C. Stokes. 2003. "Endogenous Democratization." *World Politics* 55(4): 517–549.

Bollen, Kenneth A. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 45: 370–90.

_____. 1993. "Liberal Democracy: Validity and Method Factors in Cross-National Measures." *American Journal of Political Science* 37(4): 1207–1230.

_____. 2001. "Cross-National Indicators of Liberal Democracy, 1950–1990." 2nd ICPSR version. Chapel Hill, NC: University of North Carolina and Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Bollen, Kenneth A., and Pamela Paxton. 2000. "Subjective Measures of Liberal Democracy." *Comparative Political Studies* 33(1): 58–86.

Bollen, Kenneth A., and Robert W. Jackman. 1989. "Democracy, Stability, and Dichotomies." *American Sociological Review* 54(4): 612–621.

Bowman, Kirk, Fabrice Lehoucq, and James Mahoney. 2005. "Measuring Political Democracy: Case Expertise, Data Adequacy, and Central America." *Comparative Political Studies* 38(8): 939–970.

Buhlmann, Marc, Wolfgang Merkel, and Bernhard Wessels. 2008. "The Quality of Democracy: Democracy Barometer for Established Democracies." National Centre of Competence in Research (NCCR) of the Swiss National Science Foundation, Working Paper No. 10a.

Carson, Carol S. 1975. "The History of the United States National Income and Product Accounts: The Development of an Analytical Tool." *Journal of Income and Wealth* 21: 153–181.

Casper, Gretchen, and Claudiu Tufis. 2003. "Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets." *Political Analysis* 11: 196–203.

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(1): 87–119.

Cheibub, Jose Antonio, and Jennifer Gandhi. 2004. "Classifying Political Regimes: A Six-Fold Measure of Democracies and Dictatorship." Presented at the annual meetings of the American Political Science Association, Chicago, September 2–5.

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

Cingranelli, David L., and David L. Richards. 1999. "Measuring the Pattern, Level, and Sequence of Government Respect for Human Rights." *International Studies Quarterly* 43: 407–417.

_____. n.d. The CIRI Human Rights Dataset, http://ciri.binghamton.edu/index.asp.

Clark, Terry and Vincent Hoffmann-Martinot, eds. 1998. *The New Political Culture*. Boulder: Westview Press.

Cohen, Joshua. 1989. "Deliberation and Democratic Legitimacy." In *The Good Polity*, ed. Alan Hamlin and

Phillip Petit. Oxford: Blackwell.

Collier, David, and Robert Adcock 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts." *Annual Review of Political Science* 2:537–565.

Collier, David, and Steven Levitsky. 1997. "Democracy with Adjectives: Conceptual Innovation in Comparative Research." *World Politics* 49: 430–451.

Coppedge, Michael. 1999. "Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics." *Comparative Politics* 31(4): 465–476.

Coppedge, Michael, Angel Alvarez, and Claudia Maldonado. 2008. "Two Persistent Dimensions of Democracy: Contestation and Inclusiveness." *Journal of Politics* 70(3): 335–350.

Coppedge, Michael; John Gerring; et al. 2011. "Conceptualizing and Measuring Democracy: A New Approach." *Perspectives on Politics* 9(1): 247–267.

Coppedge, Michael, and Wolfgang H. Reinicke. 1990. "Measuring Polyarchy." *Studies in Comparative International Development* 25(1): 51–72.

Dahl, Robert A. 1956. *A Preface to Democratic Theory*. Chicago: University of Chicago Press.

_____. 1982. *Dilemmas of Pluralist Democracy*. New Haven: Yale University Press.

_____. 1989. *Democracy and its Critics*. New Haven: Yale University Press.

Diamond, Larry and Leonardo Morlino, eds. 2005. *Assessing the Quality of Democracy*. Baltimore: Johns Hopkins University Press.

Dryzek, John S. 2009. "Democratization as Deliberative Capacity Building." *Comparative Political Studies* 42(11): 1379–1402.

_____. 2010. *Foundations and Frontiers of Deliberative Governance*. Oxford: Oxford University Press.

Dunn, John. 1995. *Democracy: A History*. Boston: Atlantic Monthly Press.

Economist Intelligence Unit (EIU). 2010. "Democracy Index 2010: Democracy in Retreat." Economist Intelligence Unit. (http://www.eiu.com/public/democracy_index.aspx), accessed January 25, 2011.

Elkins, Zachary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations."*American Journal of Political Science* 44(2): 287–294.

Elkins, Zachary, Tom Ginsburg, and James Melton. 2011. Comparative Constitutions Project. http://www.comparativeconstitutionsproject.org/index.htm. accessed February 1, 2011.

Elster, Jon, ed. 1998. *Deliberative Democracy*. Cambridge: Cambridge University Press.

Finkel, Steve, Anibal Pérez-Liñán, and Mitchell A. Seligson. 2007. "The Effects of U.S. Foreign Assistance on Democracy Building, 1990–2003." *World Politics* 59: 404–439.

Fishkin, James S. 1991. *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven: Yale University Press.

Ford, Henry Jones. 1898/1967. *The Rise and Growth of American Politics*. New York: Da Capo.

Foweraker, Joe, and Roman Krznaric. 2000. "Measuring Liberal Democratic Performance: An Empirical a Conceptual Critique." *Political Studies* 48(4):759–787.

Freedom House. 2007. "Methodology," Freedom in the World 2007. New York. (http://www.freedomhouse. org/template.cfm?page=351&ana_page=333&year=2007), accessed September 5, 2007.

Fung, Archon. 2005. "Deliberation before the Revolution: Toward an Ethics of Deliberative Democracy in an Unjust World." *Political Theory* 33(2): 397–419.

Fung, Archon, and Erik Olin Wright, eds. 2003. *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*. London: Verso.

Gallie, W. B. 1956. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56: 167–220.

Gasiorowski, Mark J. 1996. "An Overview of the Political Regime Change Dataset." *Comparative Political Studies* 29(4): 469–83.

Gates, Scott, Håvard Hegre, Mark P. Jones, and Håvard Strand. 2006. "Institutional Inconsistency and Political Instability: Polity Duration, 1800–2000." *American Journal of Political Science* 50(4):893–908.

Gerring, John. 2012. *Social Science Methodology: A Unified Framework, 2d ed*. Cambridge: Cambridge University Press.

Gerring, John, Philip Bond, William Barndt, and Carola Moreno. 2005. "Democracy and Growth: A Historical Perspective." *World Politics* 57(3): 323–364.

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: 361–83.

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

Goodnow, Frank J. 1900. *Politics and Administration*. New York: Macmillan.

Gordon, Scott. 1999. *Controlling the State: Constitutionalism from Ancient Athens to Today*. Cambridge: Harvard University Press.

Gutmann, Amy, and Dennis Thompson. 1996. *Democracy and Disagreement*. Cambridge: Harvard University Press.

Habermas, Jurgen. 1984. *The Theory of Communicative Action. Volume One: Reason and the Rationalization of Society*. Boston: Beacon Press.

_____. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: MIT Press.

Hadenius, Axel. 1992. *Democracy and Development*. Cambridge: Cambridge University Press.

Hadenius, Axel, and Jan Teorell. 2005. "Assessing Alternative Indices of Democracy." Committee on Concepts and Methods Working Paper Series, August.

Hamilton, Alexander, James Madison, and John Jay. 1992 [1787–88] . *The Federalist*. Ed. William R. Brock. London: Everyman/J.M. Dent.

Hayek, Friedrich A. 1960. *The Constitution of Liberty*. London: Routledge & Kegan Paul.

Held, David. 2006. *Models of Democracy*, 3d ed. Cambridge: Polity Press.

Herrera, Yoshiko M., and Devesh Kapur. 2007. "Improving Data Quality: Actors, Incentives, and Capabilities." *Political Analysis* 15(4): 365–386.

Hirst, Paul Q., ed. 1989. *The Pluralist Theory of the State: Selected Writings of G.D.H. Cole, J.N. Figgis, and H.J. Laski*. New York: Routledge.

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

IDEA. [n.d.] Global Database of Quotas for Women. (http://www.quotaproject.org/), accessed February 1, 2011. IDEA. [n.d.] Global Survey of Voter Turnout. (http://www.idea.int/vt/index.cfm), accessed February 1, 2011.

Interparliamentary Union (IPU). [n.d.] Women in National Parliaments Statistical Archive. (http://www.ipu.org/ wmn-e/classif-arc.htm) , accessed February 1, 2011.

Isaac, Jeffrey C. n.d. “Thinking About the Quality of Democracy and its Promotion.” Unpublished ms.

Jackman, Simon. 2008. “Measurement.” In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford, UK: Oxford University Press.

Kelley, Judith. 2009. “The More the Merrier? The Effects of Having Multiple International Election Monitoring Organizations.” *Perspectives on Politics* 7(1): 59–64.

Krook, Mona Lee. 2009. *Quotas for Women in Politics: Gender and Candidate Selection Reform Worldwide*. Oxford: Oxford University Press.

Kurtz, Marcus J. and Andrew Schrank. 2007. “Growth and Governance: Models, Measures, and Mechanisms.” *Journal of Politics* 69(2): 538–554.

Landman, Todd. 2003. “Map-Making and Analysis of the Main International Initiatives on Developing Indicators on Democracy and Good Governance.” Unpublished manuscript, University of Essex.

_____. 2004. “Measuring Human Rights: Principle, Practice, and Policy.” *Human Rights Quarterly* 26(4):906–931.

Landman, Todd, and Edzia Carvalho. 2009. *Measuring Human Rights*. London: Routledge.

_____. 2008. *Issues and Methods in Comparative Politics: An Introduction*, 3rd Edition. London: Routledge.

Levi, Margaret. 1996. “Social and Unsocial Capital: A Review Essay of Robert Putnam’s Making Democracy Work.” *Politics & Society* 24(1): 45–55.

Lijphart, Arend. 1999. *Patterns of Democracy*. New Haven: Yale University Press.

Lindberg, Staffan I. 2006. *Democracy and Elections in Africa*. Johns Hopkins University Press.

Lindblom, Charles E. 1977. *Politics and Markets: The World’s Political-Economic Systems*. New York: Basic Books. Lively, Jack. 1975. *Democracy*. Oxford: Basil Blackwell.

Linzer, Drew A. and Jeffrey Lewis. 2010. “poLCA: Polytomous Variable Latent Class Analysis.” R package version 1.2. http: //userwww.service.emory.edu/~dlinzer/poLCA.

Linzer, Drew A. and Jeffrey K. Staton. “A Measurement Model for Synthesizing Multiple Comparative Indicators: The Case of Judicial Independence.” Paper Prepared for presentation at the 2011 Annual Meeting of the American Political Science Association, September 1-4, Seattle, Washington.

Lowell, A. Lawrence. 1889. *Essays on Government*. Boston: Houghton Mifflin.

Macpherson, C.B. 1977. *The Life and Times of Liberal Democracy*. Oxford: Oxford University Press.

Mahoney, James, and Dietrich Rueschemeyer, eds. 2003. *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

Mansbridge, Jane. 1983. *Beyond Adversarial Democracy*. Chicago: University of Chicago Press.

Marshall, Monty G. and Keith Jaggers. 2007. “Polity IV Project: Political Regime Characteristics and Transitions, 1800–2006.” (http://www.systemicpeace.org/ inscr/p4manualv2006.pdf), accessed February 1, 2011.

McFaul, Michael. 2005. “Transitions from Postcommunism.” *Journal of Democracy* 16(3): 5–19.

McHenry, Dean E., Jr. 2000. “Quantitative Measures of Democracy in Africa: An Assessment.” *Democratization* 7(2): 168–185.

Mill, John Stuart. 1958 [1865]. *Considerations on Representative Government*, 3d ed., ed. Curren V. Shields. Indianapolis: Bobbs-Merrill.

Moon, Bruce E., Jennifer Harvey Birdsall, Sylvia Ciesluk, Lauren M. Garlett, Joshua H. Hermias, Elizabeth Mendenhall, Patrick D. Schmid, and Wai Hong Wong. 2006. “Voting Counts: Participation in the

Measurement of Democracy." *Studies in Comparative International Development* 41(2): 3–32.

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

Munck, Gerardo L., and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy: Alternative Indices." *Comparative Political Studies* 35(1): 5–34.

Mutz, Diana C. 2008. "Is Deliberative Democracy a Falsifiable Theory?" *Annual Review of Political Scienc*e 11: 521–538.

Nielson, Daniel L., Ryan M. Powers, and Michael J. Tierney. 2010. "Broad Trends in Foreign Aid: Insights from PLAID 1.9.1." Presented at the conference: Aid Transparency and Development Finance: Lessons and Insights from AidData. Oxford, March 22–25.

Nunn, Nathan. 2009. "The Importance of History for Economic Development." *Annual Review of Economics* 1(1): 1–28.

Pateman, Carole. 1976. *Participation and Democratic Theory*. Cambridge: Cambridge University Press.

Paxton, Pamela. 2000. "Women's Suffrage in the Measurement of Democracy: Problems of Operationalization." *Studies in Comparative International Development* 35(3): 92–110.

Paxton, Pamela, Jennifer Green, and Melanie Hughes. 2008. Women in Parliament, 1945–2003: CrossNational Dataset [Computer file] . ICPSR24340-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-12-22. doi:10.3886/ICPSR24340.

Paxton, Pamela, Kenneth A. Bollen, Deborah M. Lee, and Hyojoung Kim. 2003. "A Half-Century of Suffrage: New Data and a Comparative Analysis." *Studies in Comparative International Development* 38(1): 93–122.

Pemstein, Daniel, Stephen Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4): 426–449.

Persson, Torsten, and Guido Tabellini. 2006. "Democratic Capital: The Nexus of Political and Economic Change." Unpublished manuscript.

Pitkin, Hanna Fenichel. 1967. *The Concept of Representation*. Berkeley: University of California Press.

Pocock, J.G.A. 1975. *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*. Princeton: Princeton University Press.

Proyecto Estado de la Nación 2001. *Informe de la auditoría ciudadana sobre la calidad de la democracia en Costa Rica*. San José, Costa Rica: Proyecto Estado de la Nación.

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

Ranney, Austin. 1962. *The Doctrine of Responsible Party Government: Its Origins and Present State*. Urbana: University of Illinois.

Ryfe, David M. 2005. "Does Deliberative Democracy Work?" *Annual Review of Political Science* 8: 49–71.

Saward, Michael. 1998. *The Terms of Democracy*. Oxford: Polity.

_____. 2003. *Democracy*. Cambridge: Polity.

Schattschneider, E.E. 1942. *Party Government*. New York: Rinehart.

Schedler, Andreas. 2006a. "Patterns of Manipulation: Towards a Topography of Authoritarian Regimes, 1980–2002." Unpublished manuscript, Department of Political Science, CIDE.

_____. 2006b. "When Do Losers Protest?" Unpublished manuscript, Department of Political Science,

CIDE.

Schneider, Carsten Q. 2010. "Issues in Measuring Political Regimes." Working Paper 2010/12, Center for the Study of Imperfections in Democracy, Central European University. (http://disc.ceu.hu/working-papers).

Schneider, Carsten Q., and Philippe C. Schmitter. 2004a. "The Democratization Dataset, 1974–2000." Unpublished manuscript, Department of Political Science, Central European University.

_____. 2004b. "Liberalization, Transition and Consolidation: Measuring the Components of Democratization." *Democratization* 11(5): 59–90.

Schumpeter, Joseph A. 1950 [1942]. *Capitalism, Socialism and Democracy*. New York: Harper & Bros.

Shapiro, Ian. 2003. *The State of Democratic Theory*. Princeton: Princeton University Press.

Shapiro, Ian, and Casiano Hacker-Cordon, eds. 1999. *Democracy's Value*. Cambridge: Cambridge University Press.

Steiner, Jurg, Andre Bachtiger, Markus Sporndli, and Marco R. Steenbergen. 2004. *Deliberative Politics in Action: Analysing Parliamentary Discourse*. Cambridge: Cambridge University Press.

Steinmo, Sven, Kathleen Thelen, and Frank Longstreth, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Cambridge: Cambridge University Press.

Sudders, Matthew, and Joachim Nahem. 2004. *Governance Indicators: A Users' Guide*. Oslo: UNDP.

Thomas, Melissa A. 2010. "What Do the Worldwide Governance Indicators Measure?" *European Journal of Development Research* 22: 31–54.

Thompson, Dennis F. 2008. "Deliberative Democratic Theory and Empirical Political Science." *Annual Review of Political Science* 11: 497–520.

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

US Department of State. [1975–2009]. Country Reports on Human Rights Practices. Washington, DC: US House of Representatives, Committee on Foreign Affairs, and Senate, Committee on Foreign Relations.

US Office of Personnel Management Investigations Service. 2001. Citizenship Laws of the World. (http:// www.multiplecitizenship.com/documents/IS-01.pdf), accessed February 1, 2011.

USAID. 1998. Handbook of Democracy and Governance Program Indicators. Technical Publication Series PNACC-390. Washington: USAID Center for Democracy and Governance.

Vanhanen, Tatu. 2000. "A New Dataset for Measuring Democracy, 1810–1998." *Journal of Peace Research* 37: 251–265.

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

Vile, M.J.C. 1967/1998. *Constitutionalism and the Separation of Powers*. Liberty Fund.

Weale, Albert. 1999. *Democracy*. New York: St. Martin's Press.

Wilson, Woodrow. 1956 [1885]. *Congressional Government*. Baltimore: Johns Hopkins University Press.

Wu, Lawrence L. 2000. "Some Comments on 'Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect.' " *Sociological Methods and Research* 29(1): 41–64.

Table 1: Democratic Principles and Related Conceptions

I. Electoral *Aka:* elite, minimal, realist, Schumpeterian. *Principles:* contestation, competition. *Question:* are important government offices filled by free and fair multiparty elections? *Institutions:* elections, political parties, competitiveness, turnover.	**II. Liberal** *Aka:* pluralist. *Principles:* limited government, multiple veto points, horizontal accountability, individual rights, civil liberties, transparency. *Question:* Is power decentralized & constrained? *Institutions:* multiple, independent, and decentralized, with special focus on the role of the media, interest groups, the judiciary, a written constitution with explicit guarantees. *Pre-condition:* electoral democracy.
III. Majoritarian *Aka:* responsible party government. *Principles:* majority rule, centralization, vertical accountability. *Question:* Does the majority (or plurality) rule? *Institutions:* consolidated and centralized, with special focus on the role of political parties. *Pre-condition:* electoral democracy.	**IV. Consensual** *Principle:* power sharing *Question:* How diverse are the groups that participate in the formulation of public policies? *Institutions:* multiparty system, proportional electoral laws, supermajorities, oversized cabinets, federalism *Pre-condition:* electoral democracy.
V. Participatory *Principle:* government by the people. *Question:* Do ordinary citizens participate in politics? *Institutions:* election law, civil society, local government, direct democracy. *Pre-condition:* electoral democracy.	**VI. Deliberative** *Principle:* government by reason. *Question:* Are political decisions the product of public deliberation? *Institutions:* media, hearings, panels, other deliberative bodies. *Pre-condition:* electoral democracy.
VII. Egalitarian *Principle:* political equality. *Question:* Are all citizens equally empowered? *Institutions:* social and political factors that create the conditions for political equality. *Pre-condition:* electoral democracy.	

Table 2: Combining Expert Scores for an Ordinal Indicator

			Weighted Priors		
Expert	Rating (1-3)	Confidence	P(x=1)	P(x=2)	P(x=3)
1	1	50	0.14	0.05	0.02
2	2	90	0.01	0.36	0.01
3	2	95	0.01	0.39	0.01
		Total	0.16	0.80	0.04