V-Dem
VARIETIES OF DEMOCRACY

Methodology

Version 2 (September 2014)

Principal Investigators
- **Michael Coppedge** – U. of Notre Dame
- **John Gerring** – Boston University
- **Staffan I. Lindberg** – U. of Gothenburg
- **Jan Teorell** – Lund University

Research Fellows
- **Eitan Tzelgov** – U. of Gothenburg
- **Yi-ting Wang** – U. of Gothenburg

Project Managers
- **David Altman** – Pontificia U. Católica de Chile
- **Michael Bernhard** – University of Florida
- **M. Steven Fish** – UC Berkeley
- **Adam Glynn** – Emory University
- **Allen Hicken** – University of Michigan
- **Carl Henrik Knutsen – University of Oslo**
- **Matthew Kroenig** – Georgetown U.
- **Kelly McMann** – Case Western Reserve
- **Daniel Pemstein** – North Dakota State
- **Megan Reif** – U. of Gothenburg
- **Svend-Erik Skaaning** – Aarhus University
- **Jeffrey Staton** – Emory University

Institutional Homes:

UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

KELLOGG INSTITUTE

With support from:

Riksbankens Jubileumsfond

Knut och Alice Wallenbergs Stiftelse

REGERINGSKANSLIET
Ministry for Foreign Affairs Sweden

European Commission

Vetenskapsrådet

MINISTRY OF FOREIGN AFFAIRS OF DENMARK
DANIDA | INTERNATIONAL DEVELOPMENT COOPERATION

Ministry of Science, Innovation and Higher Education

Foreign Affairs, Trade and Development Canada

Forskningsrådet

QOG

IDEA

INSTITUTO DE CIENCIA POLÍTICA
FACULTAD DE HISTORIA, GEOGRAFÍA Y CIENCIA POLÍTICA

MILENIO
NÚCLEO DEMOCRACIA

Table of Contents

This document introduces the *Varieties of Democracy* (V-Dem) dataset. Part I sets forth the V-Dem conceptual scheme. Part II discusses the process of data collection. Part III describes the measurement model along with efforts to identify and correct errors.

Several additional documents associated with the V-Dem project are issued separately. *Organization and Management* introduces the project team, the web site, outreach to the international community, funding, progress to date, and sustainability. The *Codebook* includes a comprehensive list of questions and response-categories. *Country Coding Units* explains each country included in the dataset, with notes pertaining to years covered and any special circumstances that may apply (e.g., with respect to a country's pre-independence history). These documents are available from the V-Dem web site (v-dem.net).

Also on the web site is an up-to-date listing of our many collaborators, without which this project would not be possible. This includes three Project Coordinators, two post-doctoral fellows, thirty-six Regional Managers, seventeen International Advisory Board members, twenty-plus Research Assistants, and nearly two hundred Country Coordinators. (The identity of about three thousand Country Experts is not revealed so as to preserve their anonymity.) V-Dem is a global collaborative effort.

The web site also serves as the repository for other information about the project including Country Reports, working papers and publications, grant and fellowship opportunities, and the data itself. In March, 2014, a beta version of the data for 68 countries will be available for exploration in online graphs and maps. At the end of 2015 we expect to release all data collected thus far along with an updated measurement model.

1. Conceptual Scheme

There is no consensus on what democracy at-large means, beyond the vague 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).

At the same time, it is important to recognize that interpretations of democracy are not infinite. A thorough search of the literature on this protean concept reveals seven key principles: electoral, liberal, majoritarian, consensual, participatory, deliberative, and egalitarian, as summarized in Table 1. Each represents a different way of understanding what is required in order to realize "rule by the people." Thus, while no single principle can reasonably purport to embody all the meanings of democracy, these seven principles, taken together, offer a fairly comprehensive accounting of the concept of democracy as it is employed today.[1]

The *electoral* principle of democracy—sometimes referred to as contestation,

---

[1] The consensus is only relative: it could be debated if the list should consist of seven, six, or five principles, whether they should be called "principles" or "dimensions," and 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 intended to provide assurance that we are trying to take a comprehensive inventory of core values of democracy that is likely to include almost all the components that any user would want to have measured.

competition, and elite, minimal, or realist, 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 (Dahl 1956; Przeworski et al. 2000; Schumpeter 1950).

In the V-Dem conceptual scheme, electoral democracy is fundamental: we would not want to call a regime without elections "democratic" in any sense. At the same time, we recognize the many critiques of electoral democracy as being insufficient. These critiques have given rise to additional properties, each of which is designed to correct one or more limitations of electoral democracy.

The *liberal* principle of democracy addresses loopholes in electoral democracy that could allow an elected leader to act like a dictator between elections. It stresses the intrinsic importance 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.[2]

The *majoritarian* principle of democracy (aka responsible party government) reflects the principle that the will of the majority should be sovereign. According to this principle, electoral democracy is improved if it ensures that the many prevail over the few. To facilitate this, political institutions must centralize and concentrate, rather than disperse, power (within the context of competitive elections), e.g., strong and centralized parties, a unitary constitution, plurality electoral laws (or PR with high statutory thresholds), and so forth.[3]

The *consensual* (aka pluralist) principle of democracy emphasizes just the contrary: that a narrow elected majority must not be able to disregard the interests of political minorities. It favors multiple veto-points and the inclusion of as many political perspectives as possible rather than decisionmaking by narrow majorities or pluralities. Consensual democracy therefore emphasizes proportional electoral laws, large party systems, supermajority cabinets, supermajority decision rules, the separation of executive and legislative powers, a federal constitution, and other institutions that require the national head of government to share power with other political actors and bodies.[4]

The *participatory* principle of democracy is usually viewed as a lineal descendant of the "direct" (i.e., non-representative) model of democracy, as derived from the experience of

---

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

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

[4] See Lijphart (1999), Mansbridge (1983).

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 a bedrock practice of electoral democracy: delegating authority to representatives. Direct rule by citizens is preferred, wherever practicable. And within the context of representative government, the participatory element is regarded as the most democratic element of the polity. This model of democracy thus takes suffrage for granted, emphasizing turnout (actual voting) as well as non-electoral forms of participation such as citizen assemblies, party primaries, referenda, juries, social movements, public hearings, town hall meetings, and other forums of citizen engagement.[5]

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. According to this principle, democracy requires more than an aggregation of existing preferences. There should also 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 serve 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.[6]

The *egalitarian* principle of democracy stresses that formal political rights and civil liberties are insufficient for political equality. The polity should also address material and immaterial inequalities that inhibit the actual exercise of these rights and liberties. Ideally, groups – as defined by income, wealth, education, ethnicity, religion, caste, race, language, region, gender, sexual identity, or other ascriptive characteristics – should have approximately equal participation, representation, agenda-setting power, protection under the law, and influence over policymaking and policy implementation. If such equality does not already exist, the egalitarian principle requires state efforts to make the distribution of socio-economic resources, education, and health more equal so as to enhance political equality. (This principle does not entail equality of power between leaders and citizens, as leaders in all polities are by definition more powerful; but it does require efforts to move in the direction of greater equality.) [7]

---

[5] See Barber (1988), Benelo & Roussopoulos (1971), Dewey (1916), Fung & Wright (2003), Macpherson (1977), Mansbridge (1983), Pateman (1976), Rousseau (1984), Young (2000).

[6] See Bohman (1998), Elster (1998), Fishkin (1991), Gutmann & Thompson (1996), Habermas (1984, 1996), Held (2006, ch. 9), Offe (1997). 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).

[7] See Ake (1999), Berman (2006), Bernstein (1961, 1996), Dahl (1982, 1989), Dewey (1916, 1930), Dworkin (1987, 2000), Gould (1988), Lindblom (1977), Meyer (2007), Offe (1982), Sen (1999), Walzer & Miller (1995). Many of

Naturally, the conceptual scheme presented in Table 1 does not capture all the theoretical distinctions at play in the complex concept of democracy. 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; Held 2006; Shapiro and Hacker-Cordon 1999). All typologies are limited in some respects. Nevertheless, the typology 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 seven principles among the world's polities. Some countries will be particularly strong on the electoral principle; others will be strong on the egalitarian principle, and so forth. Thus, the typology provided in Table 1 is likely to prove a useful empirical device, allowing one to chart variation in political institutions through time and space.

The typology constitutes one step in the direction of disaggregation. A second step is the identification of meso-level *components*. These specify different aspects of the seven principles, and are too numerous to list here. They also exist at multiple levels; i.e., a single component may have several sub-components. (All are referred to generically as components.)

The final step in disaggregation is the identification of *indicators*, the empirical referents of components. In identifying indicators we look for features (a) that are related to at least one principle of democracy; (b) that bring the political process into closer alignment with the core meaning of democracy (rule by the people); and (c) that are measurable, directly or indirectly (via coder judgments) across polities and through time.

Indicators take the form of nominal data (classifications, text, dates) and ordinal or interval scales. Some 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. We are inclined to prefer *de facto* indicators, but we collect *de jure* indicators as well so that they can be compared.)

---

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.

Table 1:

Principles of Democracy

I. Electoral

Ideals: Contestation, competition.
Question: Are important government offices filled by free and fair multiparty elections before a broad electorate?
Institutions: Elections, political parties, competitiveness, suffrage, turnover.

II. Liberal

Ideals: Limited government, individual liberty, rule of law.
Question: Is power constrained and individual rights guaranteed?
Institutions: Civil liberties, independent bodies (media, interest groups, judiciary); constitutional constraints.

III. Majoritarian

Ideals: Majority rule, vertical accountability.
Question: Does the majority rule?
Institutions: Consolidated and centralized, with special focus on the role of political parties.

IV. Consensual

Ideal: Power sharing, multiple veto-points
Question: How numerous, independent, and diverse are the groups and institutions that participate in policymaking?
Institutions: Federalism, separate powers, PR, supermajorities, oversized cabinets, large party system.

V. Participatory

Ideal: Direct, active participation in decision-making by the people.
Question: Do citizens participate in political decision-making?
Institutions: voting, consultation, civil society, local government, direct democracy.

VI. Deliberative

Ideal: Government by reason.
Question: Are political decisions the product of public deliberation?
Institutions: Media, hearings, panels, other deliberative bodies.

VII. Egalitarian

Ideal: Equal distribution of power among citizens.
Question: Are all citizens equally empowered?
Institutions: Income, health, education, and other indicators of empowerment.

There are 400+ indicators in the V-Dem questionnaire. Each question, along with possible response-types, is listed in the *Codebook*. More will be said about coding procedures in the next section.

To summarize, the V-Dem conceptual scheme recognizes several levels of specificity:

    ● *Core concept (1)*
        ● *Principles (7)*
            ● *Components (~30)*
                ● *Indicators (400+)*

Clarifications

Several important clarifications apply to the V-Dem conceptual scheme.

First, our attempt to operationalize democracy through principles, components, and indicators does not attempt to incorporate the *causes* of democracy (except insofar as some elements of our far-flung concept might affect other elements). For example, regime-types may be affected by economic development (Epstein et al. 2006), by colonial experiences (Bernhard et al. 2004), or by attitudes and political cultures (Almond & Verba 1963/1989; Hadenius & Teorell 2005b; Welzel 2007). However, these elements are not regarded as *constitutive* of democracy.

Second, our quest to conceptualize and measure democracy should not be confused with the quest to conceptualize and measure *governance*.[8] Of course, many elements of democracy might also be considered elements of good governance. But 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 which we anticipate V-Dem will be able to shed light on. However, measures of good governance – unless they happen to dovetail with measures of democracy – must be gathered from other sources.

Third, we recognize that some indicators and components (listed in the *Codebook*) are more important in guaranteeing a polity's overall level of democracy than others. This of course depends upon one's model of democracy. The point is that 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 principle 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 majoritarian and consensus principles, which adopt contrary perspectives on most institutional components.

---

[8] Rose-Ackerman (1999), Thomas (2010). Inglehart & Welzel (2005) argue that *effective* democracy – as opposed to purely formal or institutional democracy – is linked to state capacity. A formally democratic country that has no capacity, or is ridden with corruption, is not democratic in the full sense of the term. In order to represent this very thick concept of democracy they multiply the Freedom House indices by indices of corruption (drawn from Transparency International or the World Bank), producing an index of effective democracy. See Hadenius & Teorell (2005b) and Knutsen (2010) for critical discussion.

One can easily perceive conflicts across other principles 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 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 dilemma for policymakers and academics to grapple with.

Fifth, our proposed set of principles, components, and indicators, while fairly comprehensive, is by no means entirely comprehensive. The protean nature of *democracy* resists closure; there are always potentially new principles/components/indicators that, from one perspective or another, may be associated with this essentially contested term. Moreover, some conceptions of democracy 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.

Sixth, it is important to acknowledge that principles and higher-order components, while much easier to define than *democracy*, are still resistant to authoritative conceptualization. Our objective in Table 1 is to identify the most essential and distinctive attributes associated with these concepts. 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 framework 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 modular qualities of a Lego set.

Finally, as should be obvious, this section approaches the subject from a *conceptual* angle. Elsewhere (e.g., in the *Codebook* and on the V-Dem web site), we offer indices that measure different principles and components of democracy. (When fully operationalized, a principle of democracy is referred to as a *democracy index*.)

2. 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.

History of Polities

Our principal concern is with the operation of political institutions within large and fairly well-defined political units which enjoy a modicum of sovereignty or which serve as operational units of governance (e.g., colonies of overseas empires). We shall refer to these units as polities or countries.[9]

We are not concerned merely with the present and recent past of these polities. In our view, understanding the present – not to mention the future – requires a rigorous analysis of history. The regimes that exist today, and those that will emerge tomorrow, are the product of complex processes that unfold over decades. Although regime changes are sometimes sudden, like earthquakes, these dramatic events are perhaps best understood as a combination of pent-up forces that build up over long spans of time, not simply the precipitating factors that release them. Likewise, recent work has raised the possibility that democracy's impact on policies and policy outcomes take effect over a very long period of time (Gerring et al. 2005). Arguably, short-term and long-term effects are quite different. For all these reasons, we believe that a full understanding of trends, causes, effects, and sequences of democratization depends upon historical data.[10]

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 preserved in books, articles, newspapers archives, and in living memory. Democracy is, after all, a high-profile phenomenon. Although a secretive regime may hide the true value of goods and services in the country it cannot disguise the existence of an election. And those features of an election that might prejudice the outcome towards the incumbent are difficult to obscure completely. Note that 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.

---

[9] We are not measuring democracy within very small communities (e.g., neighborhoods, school boards, municipalities, corporations), in contexts where the political community is vaguely defined (e.g., transnational movements), or on a global level (e.g., the United Nations). 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.

[10] This echoes a persistent theme presented in Capoccia and Ziblatt (2010), Teorell (2011), and in other historically grounded work (Nunn 2009; Mahoney & Rueschemeyer 2003; Pierson 2004; Steinmo, Thelen, & Longstreth 1992).

Thus, we regard the goal of historical data-gathering as essential and also realistic, even if it cannot be implemented for every indicator in the *Codebook*. Some historical indicators are better than none at all. Furthermore, if V-Dem can demonstrate empirically that the kind of indicators that can be coded for the past are highly correlated with indicators that can be coded only for the present, proxy historical indicators can be constructed. Finally, V-Dem's pilot phase and its subsequent data collection to date proves that it is possible to extend our detailed, disaggregated indicators far back in modern history, even for little-studied countries (e.g., Suriname, Myanmar, Albania) and even during years prior to formal independence (i.e., under colonial rule).

V-Dem therefore aims to gather data whenever possible back to 1900 for all territories that can claim a sovereign or semi-sovereign existence (they enjoyed a degree of autonomy at least with respect to domestic affairs) and served as the operational unit of governance (separable from other territories). The latter criterion means that they are governed differently from other territories and we might reasonably expect many of our indicators to vary across these units. Thus, in identifying political units we look for those that have the highest levels of autonomy and/or are operational units of governance. These sorts of units are referred to as "countries," even if they are not fully sovereign. This means that Eritrea is coded as an Italian colony (1900-41), a province of Italian East Africa (1936-41), a British holding administered under the terms of a UN mandate (1941-51), a federation with Ethiopia (1952-62), a territory within Ethiopia (1962-93), and an independent state (1993-). Sometimes, historical polities are coded separately because their existence is not tied clearly to a contemporary nation-state. For further details, see *Country Coding Units*.

V-Dem provides time-series codings that reflect historical changes as precisely as possible. The coding procedure allows coders to specify the exact dates (day/month/year) corresponding to changes in an 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.

At some point in the near future, we hope to be able to extend V-Dem coding back further in historical time, e.g., to 1800 (at least for sovereign units). This will enhance our knowledge of democratic transitions for a handful of countries whose process of democratization began prior to the twentieth century. It will also enhance our knowledge of the pre-democratic history of all countries, a history that may exert an enduring influence over subsequent developments in the 20th and 21st centuries. In any case, the present V-Dem database (extending from 1900 to the present) contains information relevant to most democratic transitions (however defined) and is sufficient for most purposes.

The most immediate priority however, is to extend the current dataset (1900 – 2012) forward and be in a position to conduct bi-annual or annual updates of all countries' data across the spectrum of indicators, component indices, and measures of varieties of democracy. We hope to have funds to start such a cycle in 2015, or at the latest in 2016.

Coding Types

The 400+ indicators listed in the *Codebook* fall into three main 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 Principal Investigators and Project Managers and carried out by Research Assistants connected to the project. Selected A-data are also cross-checked and vetted by V-Dem's Country Coordinators. Sources are listed in the *Codebook* (and will eventually be integrated into the on-line database). This sort of coding comprises several hundred indicators.

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 Coordinators under the supervision of Regional Managers. This sort of coding comprises several dozen indicators.

Type (C) data requires some degree of judgment about the state of affairs obtaining in a particular country at a particular point in time. This genre of indicator is coded by Country Experts – generally academics or policymakers who are nationals of and/or residents in a country, with deep knowledge of that country and of a specific substantive area. This sort of coding comprises about 200 indicators.

Type (D) data is created from (A), (B), or (C) coding. This includes cumulative indicators such as "number of presidential elections since 1900." It also includes more aggregated variables such as components and democracy indices, as described above. (The empirical embodiment of a principle of democracy is referred to as a *democracy index*.) There is no limit, in principle, to the number of higher-order variables that might be generated as part of the V-Dem project.

Expert Coding

Type (C) coding – by Country Experts – is the most difficult, since it involves judgment on the part of the coder. Accordingly, a number of steps are taken to minimize error and to gauge the degree of imprecision that remains.[11]

In order to ensure that we are able to recruit widely among potential experts, and in order to minimize confusion due to unfamiliarity with English, type-C questions are translated into five languages: Arabic, French, Portuguese, Russian, and Spanish. Approximately one-eighth of the coders so far have done their coding in a non-English version of the questionnaire.

We seek a minimum of five Country Experts to code each country-year for every indicator. The 177 C-indicators are organized into twelve surveys and four clusters, as follows:

1. Elections
   Political parties/electoral systems
   Direct democracy

2. Executive
   Legislature
   Deliberation

---

[11] For a perceptive discussion of the role of judgment in coding see Schedler (2012a).

3. Judiciary
   Civil liberty
   Sovereignty

4. Civil society organizations
   Media
   Political equality

We suggest (but do not require) that each Country Expert code at least one cluster. It is left to the Regional Managers in consultation with the coders, to decide who is most appropriate for which surveys and how many surveys a single coder will be assigned. Note that if five Country Experts coded all four indicators for each country, V-Dem would use about 1,000 coders, in total, to collect the data on C-indicators across the world. However, most experts possess the requisite expertise to code only one or two of the clusters. This means that, in practice, each country is typically coded by a dozen or so Country Experts.[12] This in turn, means that V-Dem will engage the services of some 3,000 Country Experts on its way to coding all countries in the world.

Country Experts typically code a survey cluster(s) for a single country from 1900 to the present. They are also encouraged to code multiple countries, though in practice constraints of time or expertise often prevent them from doing so. In this event, they are encouraged to perform a simpler type of *lateral* coding for a single point in time – 1 January, 2012 – and for as many countries as they feel comfortable coding (typically, a handful). The purpose of lateral coding is to overcome country-specific biases and assure cross-country equivalence in coding by making sure that coders are forced to make explicit comparisons across countries. (A small number of Country Experts work exclusively on lateral coding.) Lateral coding always includes a *regional hub* (Australia, Brazil, Egypt, India, Indonesia, Mexico, Nigeria, Russia, South Africa, United Kingdom, or United States) and a *global hub* (Australia, France, Japan, Portugal, United Kingdom, or United States) – unless the chosen regional hub is also listed as a global hub (in which case there is only one hub country). This ensures that it will be possible to link all codings together in an integrated measurement model, as discussed below.

Informed by our experience with the pilot study, the following procedure is used to identify and select Country Experts. First, we identify a list of potential coders for a country (typically 100-150 names per country). This list is compiled by Regional Managers in consultation with Country Coordinators, based on their intimate knowledge of a country. Research Assistants (located at Gothenburg or Notre Dame) also contribute to this list, using information about potential country experts gathered from the web. Other members of the project team provide additional names if and when they have country-specific expertise.

Regional Managers and Country Coordinators thus play a critical role in the data collection process. V-Dem's approach is to recruit Regional Managers who are nationals of or residents in one of the countries in each region whenever possible. Country Coordinators are

---

[12] Note that in some rare cases, small and under-studied countries (e.g., Yemen) it is 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.

almost always nationals and residents in the country to be coded. This is critical to our effort to identify and recruit Country Experts who have expertise in various sections of the survey.

For each potential coder/Country Expert we compile basic information – country of origin, current location, highest educational degree, current position, and area of expertise. We also take note of any possible biases that might affect their ability to code questions in a dispassionate manner.

The selection and enrollment of Country Experts from the initial list is handled by Regional Managers in consultation with the Principal Investigators and the Country Coordinator. Five criteria are applied to this choice.

The most important selection criterion, naturally, is expertise in the country(ies) and the section of the survey they are assigned to code. This is usually signified by an advanced degree in the social sciences, law, or history; a record of publications; and positions in civil society that establish their expertise in the chosen area (e.g. a well-known and respected journalist). Naturally, potential coders are drawn to areas of the survey that they are most familiar with, and are unlikely to agree to code topics they know little about. So, self-selection works to achieve our primary goal of matching questions in the survey with country-specific expertise.

The second criterion is origin in the country to be coded. V-Dem's goal is that a minimum of three out of five (60%) Country Experts should be nationals or permanent residents of the country they code (preferably both). Exceptions are required for a few countries where it is difficult to find in-country coders who are both qualified and independent of the governing regime. This criterion should help avoid Western/Northern biases in the coding, which can also come from self-selection biases in who makes the migration to Western/Northern universities.

The third criterion is the prospective coder's seriousness of purpose. By this, we mean a person's willingness to devote time to the project, to deliberate carefully over the questions asked in the survey, and to report their honest judgment. Sometimes, personal acquaintanceship is enough to convince a Regional Manager that a person is fit, or unfit, for the job. Sometimes, this feature becomes apparent in communications that precede the offer to work on V-Dem.[13]

The fourth criterion is impartiality. V-Dem aims to recruit coders who will answer questions in the survey in an impartial manner. This means avoiding those who might be beholden to powerful actors – by reason of coercive threats or material incentives – or who serve as spokespersons for a political party or ideological tendency (in some instances, such as North Korea, this may entail avoiding all in-country coders). Where this is difficult, or where the reality is difficult to determine, we aim to include a variety of coders who, collectively, represent an array of views and political perspectives on the country in question.

The final criterion is obtaining diversity in professional background among the coders

---

[13] This last problem is not a significant threat to coding validity in the pilot study or the first phase of the main data collection. Few individuals seem to have been motivated to conduct this arduous coding assignment for purely monetary reasons. The reason for this is that V-Dem pays very little relative to what highly qualified experts could earn for the same amount of work from other pursuits. Further strengthening this point, there seems to be no relationship between the wealth of the country and our ability to recruit coders: we have faced challenges getting experts to agree to conduct coding for the poorest as well as the richest countries in the world.

chosen for a particular country. This entails a mixture of professionals who are active in political affairs (e.g., in the media or civil society organizations) along with academics who study these topics. It also means finding experts who are located at a variety of institutions, universities and research institutes.

After weighing these five criteria, an initial set of (typically) twenty or so Country Experts is contacted. If the quota of five Country Experts per section of the survey for each country is not filled, we work down the list of potential Country Experts until the quota is obtained. Those who fail to begin or complete the survey in a reasonable time are replaced by others, following the same procedure. Coders receive a modest honorarium for their work that is proportional to the number of surveys they have completed.

A number of steps are taken to assure informed consent and confidentiality among participants. 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. Any data released to the public that includes information about experts excludes information that might be used to identify coders.

Of course, finding the right coders is only the first step toward removing bias from a survey. The more important feature of a survey is the construction of the survey itself. In crafting indicators, we have sought to construct questions whose 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, prompting quite a few revisions in the next round of surveys. The revised questions for C-coding then went through several rounds of review with the Project Managers and outside experts over the course of two years before emerging in the current and final form.

Importantly, 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 of the survey. Here, experts can comment (in prose) on any aspect of the indicators or ratings that s/he found problematic or difficult to interpret.

3. Measurement

Having discussed the process of data collection, we proceed to the task of measurement. Under this rubric, we include (a) our measurement model, (b) methods of identifying error in measurement, (c) studies of measurement error, and (d) methods of correcting error.

Measurement Model

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. These sorts of tests – standard in most social science studies – are rarely if ever employed in extant democracy indices. To combine expert ratings for a particular country/indicator/year to generate a single "best estimate" for each question, we employ methods inspired by the psychometric and educational testing literature (see e.g. Lord & Novick 1968, Jonson & Albert 1999, Junker 1999, Patz & Junker 1999). While we select experts carefully, they clearly exhibit varying levels of reliability and bias. In such circumstances, the literature recommends that researchers use latent measurement models to aggregate diverse measures where possible, incorporating information characterized by a wide variety of perspectives, biases, and levels of reliability (Bollen & Paxton 2000, Clinton & Lapinski 2006, Clinton & Lewis 2008, Jackman 2004, Treier & Jackman 2008, Pemstein, Meserve & Melton 2010). Therefore, we use Bayesian item response theory (IRT) modeling techniques (Fox 2010) to estimate latent polity characteristics from our collection of expert ratings. The underpinnings of these measurement models are straightforward: they use patterns of cross-rater (dis)agreement to estimate variations in reliability and systematic bias across disparate measures of the same, or similar, concepts (i.e. multiple expert ratings). In turn, these techniques make use of the bias and reliability estimates to adjust estimates of the latent—that is, only indirectly observed—concept (e.g. executive respect for the constitution, judicial independence, or property rights) in question.

Our latent variable models use agreement patterns to quantify the relative degree of confidence that users should place in individual ratings. Variation in these confidence estimates reflect situations where experts disagree, or where little information is available because few raters have coded a case. These confidence estimates are tremendously useful. Indeed, the tendency of most researchers to treat the quality of measures of complex, unobservable concepts as equal across space and time, ignoring dramatic differences in ease of access and measurement across cases, is fundamentally misguided, and constitutes a key threat to inference. It is implausible, for example, that estimates pertaining to the democratic quality of countries in the 1910s will be as precise as estimates of those same countries in the 1990s. Even among modern cases, access to information about regime characteristics varies substantially across space. The assumption that all observations of latent concepts are created equal undermines the credibility of existing empirical work.

We are improving on off-the-shelf latent variable modeling approaches in two primary ways. First, hierarchical IRT modeling techniques (Patz et al. 2002, Mariano & Junker 2007) allow us to incorporate a variety of secondary data into our bias, reliability, and latent-variable estimates, and to appropriately model coder bias and reliability when building aggregate indices from individual indicators. All raters complete a post-survey questionnaire that asks demographic and attitudinal questions. Coders also report personal assessments of confidence in their responses to each question. These data will allow us to tease out patterns concerning biases and reliability across different types of experts, and generally improve the quality of our estimates.

We are also taking temporal dynamics seriously. Time enters our measurement equation in three guises. First, regime traits themselves—the latent variables of interest—

follow complicated temporal patterns. Typically, regime traits are characterized by long periods of stasis punctuated by short bursts of change. Second, like institutions, rater errors are likely to be sticky. When coders make mistakes in their evaluations, those errors are likely to influence temporally proximate ratings. For example, if the PolityIV team overestimates Turkey's constraints on the executive at time t, then they are likely to produce a very similar overestimate at time t + 1. The basic intuition here is simply that a rater is likely to perceive those institutions in similar ways within contiguous periods of time. Third, coders are likely to be better at judging recent cases than they are at rating historical institutions. Building on recent work that addresses the first issue (Linzer & Staton 2012, Schnakenberg & Fariss 2013), and more generally on research that addresses questions of dynamics in item response data (Martin & Quinn 2002, Dunson 2003, Dunson 2007), we are developing dynamic IRT models that address all three of these issues (Melton, Meserve, and Pemstein N.d.). Explicitly modeling temporally correlated errors in measurement will allow us to overcome sources of bias that are likely to be endemic in existing analyses of institutional the causes and effects of political institutions. Similarly, while our lengthy panel allows us to track over-time processes we must acknowledge that historical measurement is often accompanied by heightened uncertainty and adjust our analyses to reflect this source of error.

For nominal and dichotomous variables, we provisionally 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).

Combining Expert Scores for an Ordinal Indicator

Expert	Rating (1-3)	Confidence	*Weighted Priors*		
			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

Again, both of these coder-aggregation techniques are merely the best provisional approximations of the kinds of point estimates that will be generated by a more complex measurement model that will incorporate all of the available information into the point estimates in a more rigorous way.

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. Using measures that are accompanied by estimates of confidence is an important step in guarding against this threat and forces both academics and practitioners to be more honest in their empirical assessments of the validity of theoretical arguments and the effectiveness of applied policies. We will (and downstream users should) incorporate these reliability estimates into applied analyses, reducing overconfidence in statistical conclusions (Treier & Jackman 2008). These estimates will also enable us and subsequent users to focus on only the most reliable data or to weight analyses to correct, for instance, for heteroscedasticity when our indicators are the dependent variables and reliability is correlated with a regressor. Furthermore, they will 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.

Identifying and Handling Measurement Error

A project of this nature cannot help but encounter error. This may be the product of linguistic misunderstandings (recall that most of our coders do not speak English as their first language and some take the survey in a translated form), misunderstandings about the way in which a question applies to a particular context, factual errors, errors due to the scarcity or ambiguity of the historical record, differing interpretations about the reality of a situation, coder inattention, errors introduced by the coder interface or the handling of data once it has been entered into the database, or random processes that we cannot generalize about or do not yet understand.

Some of these errors are stochastic in the sense of affecting the precision of our estimates but not their validity. Point scores, as estimated by the measurement model, are centered on the true value. Other errors are systematic, introducing bias into the estimates. It is this latter sort of error that is of principal concern (though naturally we would like to enhance precision as much as possible). Hereafter, when we use the term error we are referring primarily to systematic error.

To evaluate possible errors we employ a number of tests, some of which are incorporated into the measurement model and others of which are applied ex post to examine the validity of the measurement model. Our attention will be primarily focused on that portion of the questionnaire dealing with questions of a more subjective nature, i.e., C-type questions coded by Country Experts.

Every V-Dem coder completes a short questionnaire that delves into factors of possible relevance to coder judgments. This includes personal characteristics such as *sex*, *age*, *education* (highest degree received), *location of graduate work*, *place of birth* (in-country or abroad), *number of years spent in-country*, *number of years since last sojourn in-country*, *location of employer* (in-country or not), and *type of employer* (a 9-part classification that includes the important distinction between government employees and others).

The post-survey questionnaire also enquires into opinions that Country Experts hold about the country they are coding. Thus, we ask coders to assign a point score on a 0-100 scale summarizing the overall level of democracy obtaining on January 1, 2012, in that country, using whatever understanding of democracy they choose to apply. We ask the same question about several prominent countries from around the world that embody varying characteristics of democracy/autocracy. 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.

The post-survey questionnaire, finally, contains several questions intended to elicit the coder's views about the concept of democracy. The first question begins with the following statement: "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, along a Likert scale. Next, the coder is presented with a brief explanation of the seven principles of democracy outlined above – electoral, liberal, majoritarian, consensus, participatory, deliberative, and egalitarian. After each principle, the coder is asked to rate his/her degree of support for that principle on a Likert scale.

Each piece of information from this post-survey questionnaire will be used to help inform the measurement model, i.e., to enhance precision and limit possible biases. Unfortunately, because of the time required to construct and test the measurement model it was not possible to implement these elements in the first iteration of the dataset. (They are slated for the next iteration.)

The measurement model will also take into account information we can glean from the performance of the coders that might serve as an indication of their level of attentiveness, effort, and knowledge. This includes *inter-coder reliability* (assessed at the coder level across all codings), self-reported *confidence* (in each coding), *number of country-years coded* (all together), *coding changes* (the number of times that a coder changes their coding from T-1 to T relative to other coders for that country/indicator, aggregated across all codings), *time on task* (the number of hours a coder is logged into the on-line system, discounted by the number of country/indicator/years s/he has coded), *accesses* (the number of times the on-line survey is accessed), *contacts* (writing comments or asking questions of the V-Dem team that are non-logistical in nature), and *response rate* (assessed at the country level). (With the exception of inter-coder reliability, these elements have not yet been included in the model.)

Each of the foregoing features will also be tested independently. Thus, we will be able to report on whether, and to what extent, each of the observed and self-reported features of the coders affects their ratings. In particular, by including hierarchical priors that depend on observed rater characteristics and behavior in our latent variable model specifications—an approach often referred to as "empirical Bayes"—we can evaluate the extent to which such features help to explain rater bias and reliability, while simultaneously incorporating that information into indicator estimates.

In addition, we will apply several *ex post* tests to evaluate the quality of the data emanating from the measurement model. One sort of test relies on the distribution of the data. If the distribution of responses for a particular country/indicator/year is bi-modal we have an obvious problem: coders disagree wildly. This also means that the point estimate from the measurement model is unstable: a change of coding for any single coder, or the addition of a new coder, is likely to have a big impact on the point estimate. Disagreement as registered by a bi-modal distribution could represent a situation in which the truth is recalcitrant – presumably because available information about a topic is scarce and/or contradictory. Or it could represent errors that are corrigible.

A second approach to validation compares V-Dem components with other indices that purport to measure similar concepts, i.e., *convergent validity*. For example, 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"). Additional tests will be conducted with other indices that correspond to V-Dem components, e.g., the World Justice Project's Rule of Law Index (Agrasti et al. 2011). Unfortunately, techniques of convergent validity are limited in their utility. 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 components 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 with V-Dem components.

A third approach to validation focuses on *face validity*. Once data collection is complete for a group of countries, we look closely at point estimates with especially large confidence intervals, as assigned by the measurement model (for C type indicators). These are given special scrutiny by Regional Managers and other members of the V-Dem team, with recourse to graphs of the coder-level results.

Once the project is complete and the data made available to the general public, we hope to implement wiki blogs on the V-Dem web site so that end-users can identify possible errors. This might be directed at specific indicators, specific countries, or specific country/indicator/years. It might call attention to additional information—perhaps unknown to the coders—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, providing further information upon which to base estimates of uncertainty. This constitutes a second sort of face validity check, incorporating the wisdom of crowds.

Studies of Measurement Error

A final approach to validation probes various features of the data gathering process in order to gauge possible sources of error. This will take the form of various studies in which a particular issue is probed in an intensive fashion, probably focused on a small set of countries and employing a much larger sample of coders than we normally recruit.

One such study will focus on *coder types*. A key challenge to the validity is that data may be subject to the subjective perceptions and opinions of the chosen coders. Is it the case that a different set of coders might arrive at a very different set of answers? Features of the coders captured in our post-survey questionnaire can be tested systematically across the entire dataset, as noted. However, we cannot test the potential impact of a different kind of coder not included in our usual sample. This study therefore focuses on comparisons across different coder types, e.g., partisans, academics, civil society professionals, businesspeople, cosmopolitans (those speaking foreign languages and with travel or educational experience abroad), educated lay citizens, and less educated lay citizens. Results of this study should indicate (a) how far the consensus on coding extends (i.e., to what types of coders), (b) how much difference the background of the coder makes, (c) for what types of questions it matters, and (d) which sorts of coders have the most positive view of a country. More generally, we hope to learn more about the sensitivity of V-Dem data to our sampling of Country Experts.

A second study will focus on *country sequencing*. Does it matter if coders have considered other countries prior to coding Country A? Here, we randomize respondents into two groups. Group 1 is asked to code Country A. Several weeks later, they are asked to code a handful of countries including Country A, which they must re-code. The comparison cases should include those that are in the same region as well as a country (preferably in the same region, or with a history of colonial involvement in the region) generally regarded as highly democratic. Respondents are not reminded of their original codings for Country A and are encouraged to adjust their original coding if they feel that a more accurate assessment is possible, in light of their consideration of other countries. Group 2 repeats this procedure in reverse. That is, they first code a handful of related countries and then are asked to code Country A.

A third study will focus on *question ordering*. The V-Dem questionnaire is not randomized for several reasons. First, some questions must be asked in a particular order (later questions are activated or skipped depending upon the answers). Second, we wish to maintain a logical flow across questions and to make the flow as predictable as possible, so that inadvertent errors are minimized. Finally, we wish to maintain equivalence across surveys. However, we also wish to know whether the ordering of questions on the questionnaire affects the responses we get, and if so how. To probe this question we will carefully randomize questions within a survey (but not across surveys), without upsetting questions that are dependent upon others, and while maintaining some degree of logical flow. For example, we will reverse the order of questions that are asked first about men and next about women. But we will not distribute those questions randomly within the survey.

A fourth study will explore the quality of model-based *bias adjustment*. In particular, because coders from different countries may understand both question wordings and concepts

in different ways, two coders operating in different contexts might rate two identical cases differently from one another. A common approach to addressing this problem is to construct anchoring vignettes—short hypothetical depictions of cases—and then ask coders to evaluate vignettes in addition to real cases, and to use differences in vignette evaluations to correct for inter-personal differences in coder perceptions or understandings of concepts (King et. al. 2004; King & Wand 2007; Hopkins & King 2010). Because the vignettes are fixed, these techniques assume that differences in rater evaluations must represent differences in personal interpretation, and then subtract these differences from responses for real cases, ostensibly correcting for respondent incomparability. Similarly, given sufficient overlap in observed coding across raters, our latent variable modeling techniques can use patterns of inter-coder agreement to identify and correct for systematic differences in raters' perceptions and conceptual understandings. In other words, differences in how experts rate identical cases help to identify inter-expert variation in interpretation in much the same way that variation in ratings of fixed vignettes does. We can validate this feature of the model by comparing its performance to a vignette-based approach for controlling incomparability in survey responses. Focusing on a subset of indicators, we will recruit country-experts to rate an anchoring vignette, their own country, and some comparison countries. Then we will apply both vignette-based and measurement-model based corrections to responses to determine if they produce comparable results. An experimental component will also seek to determine if vignettes themselves alter coder behavior. In particular, we will use patterns of agreement between raters to determine if treated experts (vignette condition) produce codings that are systematically different from a control population (no vignette condition).

Correcting Error

Problems with *factual* questions (*A*- and *B*-type indicators) are corrected whenever the Principal Investigators, in consultation with the relevant Project Managers, are convinced that a better, i.e., more correct, answer is available.

Problems with *subjective* questions (*C*-type indicators) are handled with restraint. We fully expect that any question requiring judgment will elicit a range of answers, even when all coders are highly knowledgeable about a subject. A key element of the V-Dem project – setting it apart from Polity, Freedom House, and most other indices that rely on expert coding – is coder independence. Recall that each coder does his or her work in isolation from other coders and from members of the V-Dem team (apart from clarifying questions about the process). The distribution of responses across questions, across countries, and across years thus provides a vital clue into the relative certainty/uncertainty of each data point. Since a principal goal of the V-Dem project is to produce informative estimates of uncertainty we do not wish to tamper with evidence that contributes to those estimates. Arguably, the noise in the data is as informative as the signal.

Note also that one or two wayward coders are unlikely to have a strong influence on the point estimates that result from the measurement model's aggregation across five coders. This is especially the case if the wayward coders are consistently off-center (across all their codings); in this case, their weight in the measurement model will be reduced.

At the same time, there are occasions when error is consequential (in the sense of introducing bias) and remedial action is justifiable. One such occasion occurs when there is strong disagreement among coders for a particular country/indicator. A more complicated situation is faced when we have reason to believe that a majority of coders is wrong. This may be flagged by face validity or convergent validity tests, as discussed above.

Two procedures may be followed, separately or in tandem, to rectify errors: (a) re-coding by extant Country Experts and (b) fresh coding by new Country Experts. The first is especially relevant if there seems to be a misunderstanding (e.g., the question or the response-categories have been misinterpreted). In this case, we will contact the Country Experts to see if they are willing to re-code in light of further clarification from the V-Dem team. If, on the other hand, errors seem to be rooted in differences of opinion, the obvious expedient is to recruit additional coders.

4. References

Agrast, Mark David; Juan Carlos Botero; Alejandro Ponce. 2011. *Rule of Law Index 2011*. The World Justice Project.

Ake, Claude. 1999. *The Feasibility of Democracy in Africa*. Dakar: CODRESIA.

Almond, Gabriel A.; Sidney Verba. 1963/1989. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Newbury Park, CA: Sage.

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

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

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

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

Berman, Sheri. 2006. *Social Democracy and the Making of Europe's Twentieth Century*. Cambridge University Press,

Bernhard, Michael, Christopher Reenock, and Timothy Nordstrom. 2004. "The Legacy of Western Overseas Colonialism on Democratic Survival." *International Studies Quarterly* 48, 225-50.

Bernstein, Eduard. 1961. *Evolutionary Socialism: A Criticism and Affirmation*, Random House.

Bernstein, Eduard. 1996. *Selected Writings of Eduard Bernstein, 1900-1921*, Prometheus Books.

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

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

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

Capoccia, Giovanni; Daniel Ziblatt. 2010. "The Historical Turn in Democratization Studies: A New Research Agenda for Europe and Beyond." *Comparative Political Studies* 43 (August/September) 931-68.

Clinton, J. D.; David Lewis. 2007. "Expert Opinion, Agency Characteristics, and Agency

Preferences." *Political Analysis* 16(1): 3–20.

Clinton, J. D.; J.S. Lapinski. 2006. "Measuring Legislative Accomplishment, 1877-1994." *American Journal of Political Science* 50(1): 232–249.

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

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

Dahl, Robert A. 1982. *Dilemmas of Pluralist Democracy*. New Haven: Yale University Press.

Dahl, Robert A. 1989. *Democracy and its Critics*. New Haven: Yale University Press.

Dewey, John. 1916. *Democracy and Education*. New York: The Free Press.

Dewey, John. 1930. *Individualism*. Minton, Balch & Company.

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

Dryzek, John S. 2010. *Foundations and Frontiers of Deliberative Governance*. Oxford: Oxford University Press.

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

Dunson, David B. 2003. "Dynamic Latent Trait Models for Multidimensional Longitudinal Data." *Journal of the American Statistical Association* 98 (436): 555-563.

Dunson, David B. 2007. "Bayesian Methods for Latent Trait Modeling of Longitudinal Data." *Statistical Methods in Medical Research* 16(5): 399-415.

Dworkin, Ronald. 1987. *Law's Empire.* Cambridge: Harvard University Press.

Dworkin, Ronald. 2000. *Sovereign Virtue: The Theory and Practice of Equality*, Cambridge, MA: Harvard University Press.

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

Epstein, David L.; Robert Bates; Jack Goldstone; Ida Kristensen; Sharyn O'Halloran. 2006. "Democratic Transitions." *American Journal of Political Science* 50:3 (July) 551-69.

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.

Fox, Jean-Paul. 2010. *Bayesian Item Response Modeling: Theory and Applications.* New York: Springer.

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

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.

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

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.

Gould, Carol C. 1988. *Rethinking Democracy: Freedom and Social Cooperation in Politics, Economy, and Society*. Cambridge, U.K.: Cambridge 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.

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

Hadenius, Axel and Jan Teorell. 2005. "Cultural and Economic Prerequisites of Democracy: Reassessing Recent Evidence." *Studies in Comparative International Development* 39(4): 87–106.

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.

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.

Hopkins, Daniel, and Gary King. 2010. "Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability." *Public Opinion Quarterly*: 1-22.

Inglehart, Ronald and Welzel, Christian. 2005. *Modernization, Cultural Change and Democracy: The Human Development Sequence*. Cambridge: Cambridge University Press.

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

Jackman, Simon. 2004. "What Do We Learn from Graduate Admissions Committees? A Multiple Rater, Latent Variable Model, with Incomplete Discrete and Continuous Indicators." *Political Analysis* 12 (4): 400–424.

Johnson, Valen E,; James H Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Junker, B. 1999. "Some Statistical Models and Computational Methods that may be Useful for Cognitively-Relevant Assessment." Carnegie Mellon University.

King, Gary, Christopher JL Murray, Joshua A Salomon, and Ajay Tandon. 2004. Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research. *American Political Science Review* 98: 191–207.

King, Gary, and Jonathan Wand. 2007. Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes. *Political Analysis* 15: 46-66.

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

Lijphart, Arend. 1999. *Patterns of Democracy*. New Haven: Yale 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 and Jeffrey Staton. 2012. "A Measurement Model for Synthesizing Multiple Comparative Indicators: The Case of Judicial Independence." Emory University.

Lord, F. M. and Novick, M. R. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

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.

Mariano, Louis T. and Brian W. Junker. 2007. "Covariates of the Rating Process in Hierarchical Models for Multiple Ratings of Test Items." *Journal of the Educational and Behavioral Statistics* 32:287-314.

Martin, Andrew D., Kevin Quinn. 2002. "Dynamic Ideal Point Estimation vi Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10, 134-153.

Melton, James, Stephen A. Meserve, and Daniel Pemstein. N.d. "Time to Model the Rating Process: Dynamic Latent Variable Models for Regime Characteristics."

Meyer, Thomas. 2007. *The Theory of Social Democracy*. Cambridge: Polity Press

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

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

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

Offe, Claus. 1983. "Political Legitimation through Majority Rule?", Social Research 50 (4): 709-756.

Offe, Claus. 1997. "Micro-aspects of Democratic Theory: what makes for the deliberative competence of citizens?" in: Hadenius, Axel (ed.), *Democracy's Victory and Crisis* (Cambridge: Cambridge UP) 81-104.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. West Sussex: Wiley.

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

Patz, Richard J., and Brian W. Junker. 1999. "A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models." *Journal of Educational and Behavioral Statistics* 24: 146-178.

Patz, Richard J., Brian W. Junker, Matthew S. Johnson, and Louis T. Mariano. 2002. "The Hierarchical Rater Model for Rated Test Items and its Application to Large-Scale Educational Assessment Data." *Journal of Educational and Behavioral Statistics* 27: 341-384.

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.

Pierson, Paul. 2004. *Politics in Time: History, Institutions, and Social Analysis*. Princeton: Princeton University Press.

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.

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.

Rose-Ackerman, Susan. 1999. *Corruption and Government: Causes, Consequences, and Reform*. Cambridge: Cambridge University Press.

Rousseau, Jean-Jacques. 1984. *A Discourse On Inequality*. Trans. by Maurice Cranston. Middlesex, England: Penguin..

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.

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

Schedler, Andreas. 2012. "Judgment and Measurement in Political Science." *Perspectives on Politics* 10:1, 21-36.

Schnackenberg, Keith and Christopher J. Fariss. 2013. "Dynamic Patterns of Human Rights Practices." *Political Science Research and Methods,* in press.

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

Sen, Amartya. 1999. "Democracy as a Universal Value." *Journal of Democracy* 10 (July) 3-17.

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.

Teorell, Jan. 2011. "Over Time, Across Space: Reflections on the Production and Usage of

Democracy and Governance Data." *Comparative Democratization* 9:1 (February) 1, 7.

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.

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

Walzer, Michael and David Miller. 1995. *Pluralism, Justice, and Equality.* Oxford: Oxford University Press.

Welzel, Christian. 2007. "Are Levels of Democracy Affected by Mass Attitudes? Testing Attainment and Sustainment Effects on Democracy." *International Political Science Review* 28(4): 397–424.

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

Young, Iris M. 2000. *Inclusion and Democracy*. Oxford University Press, 2000.