

# Political Concepts

## Committee on Concepts and Methods Working Paper Series

2

April 2005

### Defining and Measuring Democracy

Michael Coppedge

University of Notre Dame

(coppedge.1@nd.edu)



**C&M**

The Committee on Concepts and Methods  
[www.concepts-methods.org](http://www.concepts-methods.org)



**IPSA**

International Political Science Association  
[www.ipsa.ca](http://www.ipsa.ca)



**CIDE**

Centro de Investigación y Docencia Económicas  
[www.cide.edu](http://www.cide.edu)

## **Editor**

Andreas Schedler (CIDE, Mexico City)

## **Editorial Board**

*José Antonio Cheibub*, Yale University

*Michael Coppedge*, University of Notre Dame

*John Gerring*, Boston University

*George J. Graham*, Vanderbilt University

*Russell Hardin*, New York University

*Evelyne Huber*, University of North Carolina at Chapel Hill

*James Johnson*, University of Rochester

*Gary King*, Harvard University

*Bernhard Kittel*, University of Amsterdam

*Gerardo L. Munck*, University of Southern California, Los Angeles

*Frederic C. Schaffer*, Massachusetts Institute of Technology

*Kathleen Thelen*, Northwestern University

**The C&M working paper series** are published by the Committee on Concepts and Methods (C&M), the Research Committee No. 1 of the International Political Science Association (IPSA), hosted at the Center for Economic Research and Teaching (CIDE) in Mexico City. C&M working papers are meant to share work in progress in a timely way before formal publication. Authors bear full responsibility for the content of their contributions. All rights reserved.

**The Committee on Concepts and Methods (C&M)** promotes conceptual and methodological discussion in political science. It provides a forum of debate between methodological schools who otherwise tend to conduct their deliberations on separate tables. It publishes two series of working papers: "Political Concepts" and "Political Methodology."

**Political Concepts** contains work of excellence on political concepts and political language. It seeks to include innovative contributions to concept analysis, language usage, concept operationalization, and measurement.

**Political Methodology** contains work of excellence on methods and methodology in the study of politics. It invites innovative work on fundamental questions of research design, the construction and evaluation of empirical evidence, theory building and theory testing. The series welcomes, and hopes to foster, contributions that cut across conventional methodological divides, as between quantitative and qualitative methods, or between interpretative and observational approaches.

**Submissions.** All papers are subject to review by either a member of the Editorial Board or an external reviewer. Authors interested in including their work in the C&M Series may seek initial endorsement by one editorial board member. Alternatively, they may send their paper to [workingpapers@concepts-methods.org](mailto:workingpapers@concepts-methods.org).

The C&M webpage offers full access to past working papers. It also permits readers to comment on the papers.

[www.concepts-methods.org](http://www.concepts-methods.org)

Scholars who set out to study a political phenomenon talk past one another if they define the phenomenon differently. Suppose two scholars want to understand “democracy,” but one understands “democracy” to refer to the liberal political democracies of advanced capitalist economies, while the other considers only the “people’s democracies” of communist regimes to be truly democratic. They will end up studying completely different countries and will probably come to opposite conclusions! If they try to reconcile their findings honestly, they will discover that they are not really studying the same phenomenon at all. Clear and consistent conceptualization is essential for preventing such misunderstandings.

Unfortunately, one of the most difficult challenges in studying democratization has been reaching agreement on what “democracy” is. In fact, W.B. Gallie once argued that “democracy” is one of the best examples of an “essentially contested” concept: a concept that is the focus of endless disputes that, “although not resolvable by argument of any kind, are nevertheless sustained by perfectly respectable arguments and evidence.”<sup>1</sup> Democracy is a contested concept because nearly everyone values the label, yet there are different reasonable and legitimate, yet incompatible, criteria for judging whether the label is deserved.

As though conceptualization were not enough of a challenge, empirical research requires us to carry out a second and a third step—operationalization and measurement. To operationalize a concept, we define a procedure for mapping a label or values of a variable onto observations in the real world. To measure a concept, we actually perform this operational procedure. The result is an indicator. Indicators are not necessarily numerical variables (although some are). Even a simple

classification of a country as a democracy is an indicator, even when the operation that produced that classification is not explicitly defined.

“Democracy” has been defined in hundreds of ways.<sup>2</sup> However, almost all definitions fit under one of four major types: economic, social, communitarian, or political democracy. (Table 2.1 about here.) Economic, social, and communitarian democracy tend to be defined in terms of outcomes: the equalization of wealth, income, and status, or the creation and maintenance of a feeling of belonging in a community or communities, and the promotion of participation within them. Political democracy is different, because it is almost necessarily defined by its procedures and institutions rather than its outcomes. Political democracy does not promise economic equality, social justice, or a feeling of community; whatever outcomes result from political democracy are consistent with this kind of democracy as long as the proper procedures produced them.

Procedural political democracy is itself divided into subtypes. All national states today have a form of representative democracy. Representation is so common that we tend to forget that there is an alternative: direct participatory democracy, in which voters make policy decisions themselves instead of electing representatives to decide for them. Representative democracy, in turn, can vary between a popular sovereignty tendency and more liberal versions. In a popular sovereignty democracy, the majority rules: whatever the people want becomes the law. Liberal democracy limits the power of the majority by guaranteeing some fundamental rights of individuals (and sometimes groups) and by creating constitutional checks on executive, legislative, and judicial powers. This set of types and subtypes is neither exhaustive nor universally accepted; one could make additional distinctions in the set of liberal representative democracies to distinguish consolidated democracies from transitional ones, parliamentary from

presidential democracies, unitary from federal democracies, high-quality from low-quality democracies, and so on. However, this basic typology is useful for describing how political scientists have faced the challenge of measuring democracy.

Although scholars would be better off if democracy were not an essentially contested concept, research on democratization is still possible. Research can proceed if, at a minimum, we are always very clear about what we mean by “democracy” so that we do not become entangled in semantic confusion. Beyond that, it is desirable to develop a consensus on a manageable number of types of democracy, so that at least some research projects address the same questions. But what is ultimately most important is to have concepts and indicators that are useful: ones that establish an easy and natural correspondence between the symbols in our minds and the observable features of the real political world that play important roles in causal processes.

Scholars make choices whenever they define or measure a concept, and their choices have consequences for their research findings. This chapter surveys the range of possible choices regarding conceptualization, levels of measurement, procedures to ensure reliability, and the aggregation of dimensions, and uses democracy indicators to illustrate the consequences of these choices.

### **Operationalizing Concepts**

In a perfect world, comparativists would use concepts that reflect the uniqueness of each country and yet are simple enough to be relevant and measurable in every country. In practice, the difficulty of gathering political information usually prevents us from achieving both goals, so we tend to settle for concepts that are either "thick" or "thin." Thick concepts have many facets; that is, they refer to many aspects of what we observe. Thin concepts have few facets: they focus

attention on only one or a few characteristics. Conceptual thickness is relative and can be understood as a matter of degree. Even a relatively thin version of democracy, one of the thickest concepts in political science, can refer to half a dozen characteristics. A thick version can refer to dozens. For example, David Held's *Models of Democracy* defines 12 different models of democracy, all of which, he argues, possess some claim to the democratic label. Between them, these 12 models refer to 72 different characteristics, which are listed in Table 2.2.

Definitions of regimes are typically thick. A good example is Juan Linz's definition of an authoritarian regime:

[Political systems without] free competition between leaders to validate at regular intervals by nonviolent means their claim to rule. . .<sup>3</sup> with limited, not responsible, political pluralism; without elaborate and guiding ideology, but with distinctive mentalities; without extensive nor intensive political mobilization, except at some points in their development; and in which a leader or occasionally a small group exercises power within formally ill-defined limits but actually quite predictable ones.<sup>4</sup>

Compare this with one set of criteria for a threshold on a democracy-nondemocracy continuum that corresponds closely to authoritarianism. I have chosen the Polyarchy Scale for this purpose because its criteria are explicitly stated. (These coding criteria are reproduced in Table 2.3.) The first two components of each definition are nearly interchangeable even though the Polyarchy Scale is more explicit here about what "limited pluralism" means in practice. (Obviously, Linz's legendary 237-page essay is far more elaborate than the brief definition quoted in Table 2.3.) The Polyarchy Scale, however, omits three additional components that are included in Linz's definition--the nature of the leaders' belief systems, the absence of active political mobilization by

the regime, and some degree of institutionalization. The Polyarchy Scale is therefore thinner.

*A Tradeoff between Concept Validity and Extension*

Operational definitions of concepts are *valid* to the extent that they refer to all the aspects of the concept that we have in mind when we use it, and no aspects that we do not have in mind. This implies that thicker concepts are not necessarily more valid. Adding a criterion to a definition makes it more valid only if the new criterion is a relevant one. For example, In the 1960s a series of scholars made the mistake of considering countries more democratic if they maintained democracy over a long period of time.<sup>5</sup> This practice only confounded democracy with stability, resulting in a loss of conceptual validity.

There is a tradeoff between the validity of a concept and its applicability to a variety of countries. Comparative research on democratization employs some thick concepts of democracy that are deeply and richly descriptive of some countries but not others, and some thin concepts of democracy that describe many countries equally well, but less revealingly. As Sartori explained, there is a tradeoff between a concept's "intension" (the number of objects to which it refers) and its "extension" (the range of countries to which can be applied).<sup>6</sup> We often say that thin concepts "travel" farther. Thick concepts do not travel as far because they carry more baggage, but they are better equipped for the places to which they do travel.

Figure 2.1 illustrates this tradeoff in extension using Linz's definitions of the basic democratic, authoritarian, and totalitarian regimes. Linz contrasted each regime with reference to five characteristics: the selection of leaders through elections, the degree of pluralism, the nature of participation, the ideological mindset of the leaders, and the degree to which the political system was institutionalized. The figure simplifies his scheme a bit by allowing each

characteristic to have only two or three possible variations. This conceptual scheme tells us a great deal about the regimes that match these characteristics. But at the same time, as the figure illustrates, the multiple requirements for each regime type limit the applicability of his definitions to just three of the 108 theoretically possible combinations.<sup>7</sup> To be more realistic about the severity of the problem, I have shaded the cells that are unlikely to contain any countries dark gray; the largest number of countries would fall in the white and light gray cells. This shading also helps illustrate the strength and weakness of thin concepts. A slightly thinner conceptual scheme that distinguished among democracy, authoritarianism, and totalitarianism based simply on elections, pluralism, and participation would probably cover all the cases in the white or light gray cells. However, in order to do so, it would tell us nothing about the omitted characteristics--institutionalization and the leaders' ideological mindset.

The ability of concepts to "travel" applies to travel in time as well as space. The ancient Athenians understood their own democracy in terms that are alien to us today. Athenians restricted democratic rights to a small minority of the adult males; they voted directly on laws rather than for representatives; they considered democracy impossible in states with more than a few thousand citizens; and they considered public-spirited harmony, not a competition among interests, essential to the nature of democracy.<sup>8</sup> By these criteria, there are no democracies in the 21st century. Today's emphasis on liberal representative democracy is the result of an eighteenth-century adaptation of the concept to the rise of large states and the struggle against absolutist monarchy. In the nineteenth century, writers (Tocqueville and Marx, for example) considered social and economic equality a defining characteristic of democracy. Most U.S. political scientists turned their backs on this tradition in the twentieth century when a purely political, procedural

version of democracy more effectively distinguished the West from its fascist and communist rivals.<sup>9</sup> Over the centuries, therefore, the concept of democracy has lost component after component, leaving us with today's very thin, "minimalist" standard that many diverse countries can satisfy.

Thin concepts of democracy are well represented by Robert Dahl's concept of polyarchy. Polyarchy has eight components, or "institutional requirements":<sup>10</sup> 1) Almost all adult citizens have the right to vote; 2) almost all adult citizens are eligible for public office; 3) political leaders have the right to compete for votes; 4) elections are free and fair; 5) all citizens are free to form and join political parties and other organizations; 6) all citizens are free to express themselves on all political issues; 7) diverse sources of information about politics exist and are protected by law; and 8) government policies depend on votes and other expressions of preference. This version of democracy is not accepted by all comparativists, but it is a well-known point of reference, and many scholars who do not wish to write their own definitions have cited polyarchy as what they mean by "democracy." With eight components, polyarchy is not the thinnest concept in the discipline, but it is thin enough to omit mention of many qualities that are commonly associated with democracy, such as majority rule, judicial independence, separation of powers, local autonomy, jury trials, and numerous personal rights, not to mention socioeconomic equality, direct democracy, small population, and public-spirited harmony.

One consequence of using a minimalist concept of democracy is that many countries qualify as "democracies" even though some subjectively seem more democratic than others. For example, in 2001-2002 Freedom House assigned its highest rating to countries as diverse as Switzerland, Uruguay, Greek Cyprus, Finland, Grenada, and the Dominican Republic.<sup>11</sup> Such

results have stimulated a reaction against minimalist concepts of democracy. Some scholars therefore remind us of components of democracy that have been dropped or taken for granted in the past 50 years and quite understandably call for them to be restored or made explicit. Thus Schmitter and Karl include institutionalization and a viable civil society ("cooperation and deliberation via autonomous group activity") among their criteria for "what democracy is."<sup>12</sup> Similarly, others stress the centrality of the rule of law and an independent judiciary.<sup>13</sup> Valenzuela and others also argue that democracy requires elected officials to enjoy autonomy from unelected "veto groups," whether they are economic conglomerates, international powers, or the military; and impartial respect for basic citizenship rights.<sup>14</sup> Guillermo O'Donnell, building on Amartya Sen, has come full circle by arguing that satisfaction of some basic economic and social needs is necessary for any meaningful democracy to exist.<sup>15</sup>

The choice of thick or thin concepts also affects the potential theoretical significance of research. Thick concepts are often meaningful only when embedded in a well-defined theory; many of them contain elaborate theoretical assumptions as elements of their definitions. They are shorthand for theories or parts of theories. Thin concepts are more theoretically adaptable: they lend themselves more easily for use in diverse theories. Philosophers of science like to remind us that all concepts are theoretical, as all constructs require making assumptions about pieces of reality that we imagine to be especially relevant for certain descriptive or explanatory purposes.<sup>16</sup> But some concepts are more theoretically involved than others.

A good way to appreciate the difference is to think of theory in the social sciences as selective storytelling. As social scientists, we craft stylized accounts of events. The elements we emphasize are the elements of theater and fiction: who the relevant actors are, what the time and

the place is (the setting), which instruments (props) can be used by the actors, the nature of their preferences or goals (motives), how they strategize to achieve their goals (plot), and a process (action) leading to a particular outcome (denouement). The thinnest concepts refer only to individual elements of a story; thick concepts tend to link together several elements. Thick concepts can be stories in themselves, sometimes complete with morals. "Dependency" was one.<sup>17</sup> Guillermo O'Donnell has formulated a series of others--bureaucratic-authoritarianism,<sup>18</sup> delegative democracy,<sup>19</sup> horizontal accountability.<sup>20</sup> The Colliers' "mode of incorporation" is yet another.<sup>21</sup> Some thick concepts would qualify as "conflicting imperatives," Andrew Gould's term for complex concepts possessing a tension that can be used to generate hypotheses.<sup>22</sup> All of these could be considered either very thick concepts or shorthand for theories.<sup>23</sup>

Unfortunately, differing preferences for thin or thick concepts lead scholars to talk past one another: when qualitative and quantitative analysts say "democracy," they literally mean different things. Strictly speaking, research on the causes of thin democracy speaks only to other research on thin democratization; research on the causes of thick democracy has relevance for a longer and richer theoretical tradition. As we shall see below, the practical consequences of defining democracy differently are not serious when the elements of different dimensions are all strongly correlated with a common underlying dimension. But sometimes changing the criteria for democracy makes a huge difference. For example, Bollen and Paxton have shown that using women's suffrage as an essential criterion for democracy changes the age of some "democratic regimes" by more than a century.<sup>24</sup>

It is important to bear in mind that the tradeoffs between thick and thin concepts are consequences of the difficulty and expense of gathering political information. If political scientists

had more resources, we could probably develop concepts of "democraticness" that would be richly descriptive of all countries (and all historical periods) to an equal degree. For the present, however, we have to choose between thick and thin concepts, and these choices have repercussions for theoretical understanding and measurement. Choices about measurement, in turn, affect the kinds of causal analysis that can be carried out. These are profoundly pivotal choices that have caused the qualitative and quantitative approaches in political science to diverge. Scholars who prefer thick concepts tend to engage in qualitative research, while those who are comfortable with thin concepts tend to promote quantitative research, and these approaches have evolved on partially separate tracks. Although these approaches have never completely lost touch with each other, it is increasingly difficult to perceive how each is relevant for the other. One of the purposes of this book is to clarify their mutual relevance.

### **Measurement**

The choice that comparativists make between thick and thin concepts affects the number of dimensions that underlie their measurements. Thick concepts tend to be multidimensional, while thin concepts tend to be unidimensional. When a concept is unidimensional, its components vary together. Intuitively, this means that if component A is present to a high degree, then component B is present to a high degree as well, and vice versa. In bivariate tables and scatterplots, unidimensional components show a strong diagonal relationship, but multidimensional components show a more even dispersion of cases in all four quadrants. Intuitively it is easy to imagine low-high or high-low combinations of multidimensional components that would not be rare exceptions. In a 2 X 2 table, cases are spread out among at least three of the four cells; in a scatterplot, they form no diagonal pattern. There is no way to

represent such patterns faithfully without employing at least two dimensions; attempting to do so would be oversimplification, or reductionism. But the higher the degree of association, the more reasonable it is to reduce the two components to one simpler concept or a single dimension.

Even minimalist concepts of democracy are usually multidimensional. Dahl, for example, argued explicitly that polyarchy had two dimensions: contestation (“the extent of permissible opposition, public contestation, or political competition”) and inclusiveness (“the proportion of the population entitled to participate on a more or less equal plane in controlling and contesting the conduct of government”).<sup>25</sup> They are separate dimensions because countries that have high contestation are not necessarily highly inclusive, and vice versa. Rather, all combinations can be observed: closed hegemonies, competitive oligarchies, inclusive hegemonies, and polyarchies. Subsequent empirical research has confirmed Dahl’s intuition. Figure 2.3 is a scatterplot of Bollen’s Political Democracy Index for 1950 (DEM50) against the percentage of the population over 20 years of age who had the right to vote in national elections in 1950 (SUFF50). DEM50 is a good indicator of contestation and SUFF50 measures the most important aspect of inclusiveness. The plot shows that, as one would expect, in 1950 there were countries with high contestation and high suffrage and countries with low contestation and no suffrage; but there were also countries with high contestation and suffrage below 50 percent, and countries with full adult suffrage but very little contestation. Just as the points in the plot cannot be joined by a straight line, the information in these two indicators cannot be reduced to a single meaningful number for each country. Contestation and suffrage therefore lie on two different dimensions.<sup>26</sup>

A thicker version of democracy would have more than two dimensions. I suspect that a thicker concept of democracy would possess five dimensions. The first two would be thick

versions of Dahl's dimensions of polyarchy--contestation (or "competition") and inclusiveness. There is probably more to contestation than becoming informed and making a simple choice among parties or candidates every few years. Contestation could also depend on the number and quality of choices presented on a ballot, democratic selection of candidates, certain kinds of public campaign financing, guaranteed media access for all parties, and opportunities for opposition parties to gain a foothold at lower levels of government.

Similarly, inclusiveness--the proportion of the adult citizens who have effective opportunities to participate equally in the available opportunities for decisionmaking--need not be confined to voting for representatives and running for office. In reality there are, or could be, many other opportunities for citizens to participate equally in decisionmaking: in judicial proceedings, at public hearings, in primaries, in referendums and plebiscites, and in speaking through the media to place issues on the public agenda. Most civil liberties fit into this dimension as well, as they involve individuals' equal right to determine their own beliefs and many other aspects of their personal lives. If the judicial system does not provide equal protection under the law, for example, the political system should be considered less inclusive. To complicate matters, inclusiveness itself may consist of two dimensions--the proportion of people possessing a right and the degree to which they possess it--which together would define a distribution of rights akin to a distribution of wealth.

To these three dimensions--contestation, breadth of inclusion, and fullness of inclusion--I would add two more: the division of powers and the scope of democratic authority. The division of powers corresponds to the unitary-federal dimension of Lijphart's concept of consensual democracy. Lijphart has established that federalism, regional autonomy, bicameralism, and local

self-government cohere as one dimension and that this dimension is distinct from his "executives-parties" dimension, which corresponds well to contestation.<sup>27</sup> Whether one considers a division of powers more democratic or merely differently democratic than unitary government is a matter of opinion, but the separateness of this dimension is beyond dispute.

A fifth dimension--the scope of democratic authority--reflects the agenda of issues that the democratic government may decide without consulting unelected actors. This dimension reflects any constraints on governmental authority imposed by the military, business groups, religious authorities, foreign powers, or international organizations regarding issues of importance to them. A broad scope of democratic authority also requires that civil servants be willing and able to implement the policies made by elected officials, because it does not matter how a government was chosen if it has no power to carry out its decisions. The fewer the issues that are in practice "off limits" to final decisionmaking by relatively inclusive bodies, the broader the scope of democratic authority. These five dimensions taken together would define democracy as a regime in which a large proportion of the citizens have an equal and effective chance to participate in making final decisions on a full range of issues at an appropriate level of government.<sup>28</sup>

Ultimately, however, the number of dimensions in a concept is an empirical question. Sometimes components that seem to be conceptually distinct are empirically associated closely enough that one can treat them as unidimensional. This is the case with the many items that are often used to measure contestation, which include regular competitive and fair elections, party competition, freedom to form and join parties and other political organizations, freedom to express diverse political positions private and publicly, and freedom for newspapers and broadcast media to express diverse points of view, especially those critical of the government. Table 2.4

provides an example of such a close association using an indicator of pluralism in the media and an indicator of freedom of expression. Most countries are arrayed along the diagonal running from the upper left to the lower right. This means that regimes that permit free expression also have laws to protect diverse media; those that forbid dissent also tend to have official media that present only the state's versions of the news; and so on. Because of this close association, both indicators can be treated as measuring the same underlying dimension, contestation. The Polyarchy Scale is a good general example of unidimensionality.<sup>29</sup> All four of its components--indicators of fair elections, freedom of organization, freedom of expression, and pluralism in the media--are closely associated. For instance, it happens that almost all countries that have many alternatives to official information also have leaders chosen in fair elections and a high degree of freedom of organization and expression; while countries in which citizens are afraid to criticize the government even privately also tend not to have meaningful elections, do not permit opposition parties or other organizations, and maintain tight official control over the media. Because of these empirical associations, it makes sense to treat these four components as reflections of a single underlying dimension, which can be called contestation.

Such unidimensionality is the reason that many democracy indicators are highly correlated: most of the aspects of democracy that they attempt to measure are the unidimensional ones related to contestation. Table 2.5 reports, for a select group of democracy indicators based on 1985 information, the average correlation of each indicator with the others. (A correlation of 1.0 would mean perfect positive association; a correlation of 0 would mean no association at all.) All of these averages are .793 or better, which is very high. It is no accident that many of these indicators have to do with aspects of contestation--"multi-party elections," "print media control,"

“freedom of political opposition,” “party legitimacy,” “competitiveness,” “right of assembly,” “broadcast media control,” etc. When so many indicators agree so closely, it is hard to escape the conclusion that we have indicators that do a pretty good job of measuring the contestation dimension of democracy.

But how accurately do they measure contestation? In measurement theory, accuracy has two components— validity and reliability. Validity—the extent to which an indicator measures what one claims it measures—has already been discussed, and it seems clear that these indicators are valid as long as we treat them as indicators of contestation rather than of democracy in all its aspects. Reliability is the degree to which a measurement procedure produces the same measurements every time, regardless of who is performing it. Reliability depends on many qualities of the measurement procedure: the reliability of the source information, the clarity of coding criteria, the skill and care of the coders, the degree of agreement among coders, and so on. But reliability is also a function of the unidimensionality of the components of an indicator because, in practice, unidimensionality is a matter of degree. The looser the associations among any of the components of an indicator, the less reliable the indicator is.

Existing democracy indicators are very reliable for identifying large differences in democracy but less reliable for measuring intermediate values and small differences. All the indicators we have can easily distinguish between Sweden and China, or even between Costa Rica and Pakistan; but none can very reliably distinguish the degrees of democracy in Greece and India in the 1990s, or the small change in Mexico from 1994 to 1997. These indicators are very useful for research on democratization that uses large samples of nearly global extension, but less useful for comparing the quality of democracy within fairly homogeneous world regions or tracking

short-term changes in single countries.

The limitations of existing democracy indicators are the result of the multidimensionality of democracy. Multidimensionality forces scholars to choose between two measurement strategies. One option is to create distinct indicators for distinct dimensions. The other is to combine all the dimensions into a single indicator. Combining dimensions is much harder to do well, so those who measure democracy have usually taken the easier path of reducing democracy to the most unidimensional set of indicators: contestation. Nevertheless, here and there scholars have created variables to measure other dimensions of democratization. Kenneth Bollen has created a sophisticated suffrage time-series; Munck and Verkuilen have created an appealing indicator of the relative strength of elected and unelected powers in Latin America; Arend Lijphart has constructed a valid indicator of the division of powers in 36 countries; and the World Bank and Transparency International have built datasets containing many indicators of corruption, bureaucratic efficiency, and other items that would be relevant for measuring the chance that the state will implement democratic decisions faithfully.<sup>30</sup> All these indicators are relevant for measuring democracy, broadly defined, but they are not integrated into a single comprehensive indicator of democracy. This strategy has the advantage of avoiding any assumptions about how these dimensions might combine to determine a country's degree of democracy. The disadvantage is that this strategy stops short of producing a single summary indicator of democracy. Paradoxically, therefore, one way to measure democracy better is to stop measuring democracy and simply measure its component dimensions instead.

This disaggregated strategy has the additional advantage of making it possible to explore empirically the interrelationships among dimensions, which would open up a fascinating new

avenue for research. Do elected officials enjoy greater autonomy vis-a-vis the military when they are backed by a broad electoral base of support? Does federalism really allow citizens to be better represented on certain issues? Does possession of the suffrage translate into effective possession of other civil and political rights? All of these are questions that should be addressed by empirical research. Such questions must be answered before any unified indicator of democracy can be developed, and it would be desirable for the answers to come from empirical research rather than mere assumptions.

The development of separate indicators is, in fact, a prerequisite for the second option: appropriate aggregation of dimensions into a single indicator of democracy. Doing this requires a stronger theory about how dimensions of democracy combine, from which one might derive a mathematical formula. Munck and Verkuilen make some suggestive remarks about aggregation rules: correspondences between certain logical relationships and certain mathematical operations.<sup>31</sup> But I suspect that a workable rule is likely to be more complex than addition and subtraction. If so, component indicators will have to be interval, if not ratio, data; otherwise, it would not be legitimate to subject them to multiplication or division, not to mention logging or exponentials.<sup>32</sup> Most measurement of democracy now is ordinal, so if we wish to develop a single indicator of democracy in several dimensions, we will have to find ways of measuring dimensions at the interval level or higher. One way to do this is to reformulate the attributes of democracy in terms of probabilities. This would entail measuring, for example, the probability that a citizen will be allowed to vote; that votes will be counted fairly; that a writer can criticize the government without being punished; and so on. These probabilities could be either estimated reasonably or calculated from actual practices. The rules for aggregating probability data are then relatively

straightforward.

A few scholars have taken steps in this direction. Axel Hadenius made a start by combining indicators of contestation and participation in an innovative and promising fashion (Hadenius 1992). Hadenius's index of democracy is an average of an indicator of open, correct, and effective elections and an indicator of various freedoms. What makes this index interesting is that before the elections component is averaged, it is multiplied by the proportion of the population that is eligible to vote and the proportion of national legislative seats that is filled by election. This mathematical operation implemented Hadenius's theoretical assumption that freedoms contribute to democracy independently of elections and that elections matter for democracy only to the extent that they select real decisionmakers and all citizens are eligible to vote. This is the kind of theory that is necessary for aggregating dimensions. However, it is not the only possible theory for doing so. Kenneth Bollen made different assumptions when combining contestation indicators with suffrage in his Liberal Democracy Series. Bollen's formula, though complex, had the effect of giving much lower scores to countries that allow competition but restrict the suffrage to about a quarter of the population or less.<sup>33</sup>

I believe that it would be useful to think in terms of a "floor" and a "ceiling" for democracy. Fundamental civil liberties constitute a floor for democracy in the sense that the freedom of individual citizens to speak, write, read, associate, and so on is valuable even if they are not allowed to compete in elections or choose representatives to make policy. A regime cannot be less democratic than the individual freedoms it allows. By a similar logic, the state's willingness and capacity to execute policies faithfully constitutes a ceiling. No matter how representative and democratic a government is, if its policies are ignored and undermined by the

bureaucracy, the police, and the courts, the whole representative process comes to nothing.

Therefore, a regime cannot be more democratic than its actual execution of any policies that are adopted democratically. Between the floor and the ceiling, what matters for democracy are all the institutions and processes designed to translate the will of the people into public policy—party competition, elections, electoral systems, legislative procedures, and executive-legislative relations.<sup>34</sup>

Unfortunately, there is as yet no scholarly consensus on a thicker definition that convincingly incorporates components such as the rule of law, the autonomy of elected officials, decentralization, or national sovereignty. Progress toward consensus would be aided by empirical analysis of the number and nature of any dimensions that structure these concepts or components. Empirical analysis is crucial because the number and nature of dimensions in a thick concept is determined more by the real world than by our imaginations. In theory, every facet of a concept could lie on a separate dimension from every other facet. In theory, for example, there could be cases in every cell of Figure 2.1: even poorly institutionalized regimes with highly ideological leaders who welcome participation and permit fair elections, but practice monistic control. It is only in practice that such combinations become odd and rare and other combinations become more common. We do not always know the reason for this. They may cause each other, or they may have a common historical cause. In any case, the dimensions that structure a thick concept are best thought of as handy bundles of a larger number of potential dimensions. Such bundles probably hold together only for selected periods and places. The more diverse the sample, and the longer the expanse of time it covers, the more likely it is to resist reduction to a small number of dimensions.

### **Consequences for Analysis**

The choices that scholars make about how to define and measure their concepts have consequences for the kinds of analysis that are possible and desirable. This section examines consequences for the selection of cases and models, for descriptive and causal inferences, and for levels of measurement.

#### *Data-Driven Research*

In comparative politics, data are scarce because the costs of collecting data are high, especially quantitative data that cover a large number of countries over a long span of time using consistent measurement criteria. The practical consequence of scarce data is that comparativists who would like to do large-sample (“large-N”), quantitative studies inevitably run up against severe constraints. They find that the variables they want to use are simply not available for many of the cases they would like to study, or that some of the variables they would like to use do not exist for any of the cases in their study, or both. If this does not lead them to give up on the project altogether, they may choose to do research on just the cases for which all their variables are available, or they may choose to drop some of their variables in order to keep a larger sample. The result is research that is “data-driven”: the choice of indicators influences the selection of cases and the set of hypotheses to be tested.

Research on democratization has always been heavily data-driven. For example, almost every quantitative study of democratization excludes some cases because of missing data, whether they are communist countries, countries with small populations, countries in civil war, or countries too far back in time for good data to exist. Another example of data-driven research is the prevalence of purely cross-sectional analyses before time-series indicators of democracy

became generally available in the late 1980s. Before then, researchers settled for data that measured democracy in a large number of countries for one year—“snapshots” of democratization frozen at one moment in time.<sup>35</sup> There was no good methodological reason to do this; in fact, we shall see in chapters 5, 7, and 11 that there are good reasons to prefer comparisons over time. In fact scholars would have preferred time-series data all along. When international relations scholars developed the Polity time series and Freedom House country ratings covered a sufficiently long span of years, time-series analysis quickly became *de rigueur* for quantitative research on democratization. Another characteristic of data scarcity has been the widespread use of indicators of debatable validity and reliability. Example abound: studies have used per capita energy consumption as a proxy for per capita gross domestic product; measured income inequality with mixed individual- and household-based data; and employed regional dummy variables as proxies for culture or world-system position. To their credit, democratization researchers have always eagerly utilized better variables and more cases as soon as they have become available; but in the meantime, the scarcity of data has always constrained their choices.

#### *Consequences of Measurement Error*

How do measurement problems affect research findings? One would expect conclusions based on somewhat unreliable indicators to inspire less confidence, but the consequences of measurement problems are not so simple. The nature of the consequences depends on whether one is trying to describe or explain; whether the measurement error is systematic or random; and, if the error is systematic, what the pattern of error is.

Description and explanation are fundamentally different tasks. Description focuses on one characteristic or variable at a time, while explanation focuses on the relationships among two or

more variables. When we are describing, that is, reporting measurements on one variable at a time, there are two kinds of errors we can commit: *systematic* or *random*. Random errors have no pattern to them; the kind or degree of error made in one measurement has nothing to do with errors made in other measurements on the same variable. If we rate some countries too high and others too low and there is no particular reason for our mistakes, then we have created random measurement error. When errors are systematic, there are reasons for our mistakes, even if we are unaware of them. We may be too tough and therefore classify some democracies as dictatorships, or we may be ethnocentric and rate presidential democracies higher than otherwise-similar parliamentary democracies. Systematic measurement error leads to biased descriptions, which are “off” in a systematic way. Random measurement error leads to “inefficient” or fuzzy descriptions, which may be unbiased on average, but are less certain. Figure 2.3 illustrates this difference by contrasting the positions of little hashmarks, representing a set of measurements on one variable, with a heavier hashmark, which shows the true value that we are attempting to measure and therefore describe. In part a, the biased measurements are to the right of the true value, showing how systematic error can lead to an overrating. When there is random error (part b), the measurements are dispersed more broadly around the true value, even though the average measurement is very close to the true value. Of course, measurement error can be biased and inefficient at the same time. All comparative indicators are probably biased and inefficient to some degree; the question is whether they are too biased or inefficient to be useful.

Democracy indicators certainly contain some measurement error, but is it random or systematic? A number of scholars have raised questions about the specific ratings of specific countries. Scott Mainwaring has questioned Przeworski et al.’s decision to code some

authoritarian regimes as democracies if they eventually surrendered power after an electoral loss, such as the military regime in Brazil from 1979 to 1984. This is an example of systematic error leading to bias. In fact, Przeworski et al. explicitly acknowledge and defend the systematic measurement error in their indicator on the grounds that it is known and correctable.<sup>36</sup> Freedom House indicators have also been criticized for some questionable ratings. For example, Scott Mainwaring has argued that Freedom House was too harsh on Latin American leftist governments in the 1980s and seems to have used stricter criteria in the 1990s than before, with the result that its ratings fail to reflect improvement in Mexico, Colombia, the Dominican Republic, El Salvador, and Guatemala in these years. Jonathan Hartlyn identifies the same discrepancies by correlating Freedom House and Polity scores for these countries.<sup>37</sup> Such discrepancies reinforce the conclusion that such indicators less reliable for small, intra-regional differences and changes than they are for large, cross-regional comparisons. If these errors are random, this is all the caution that is necessary.

However, the only way to know whether there is systematic measurement error is to analyze many measurements systematically. Two of the most sophisticated such analyses did find systematic error in several common indicators in the 1970s and 1980s.<sup>38</sup> The Banks data tended to be more favorable to Eastern Europe or communist countries and harsher on countries with recent coups, while Freedom House tended to overrate Catholic monarchies and underrate communist regimes. However, the degree of systematic error was, with one exception, 22 percent or less. Five of the eight indicators were at least 70 percent valid, and two--Freedom House's Index of Political Rights and Banks's Freedom of Group Opposition--were more than 87 percent valid.<sup>39</sup> This is fairly reassuring. Furthermore, the high correlations between this index and the others reported in

Table 2.5 suggest that even if none of these indicators is very reliable for small-N studies, quite a few of them are sufficiently reliable for large-N comparisons.<sup>40</sup>

The consequences of measurement error for causal inferences are surprising, in principle. There are three basic patterns of error: 1) the addition of a constant to either variable, 2) random error in the dependent variable, or 3) random error in the explanatory variable. Figure 2.3 illustrates these three situations using black dots for the points measured without error, white dots for points measured with error, a dashed line for the true regression line, and a solid line for the regression line affected by measurement error.

1) As King, Keohane, and Verba have pointed out, causal inferences are not affected in any way if the dependent or independent variable is systematically too high or too low.<sup>41</sup> Part c of figure 2.3 illustrates this result: when a constant bias is added (or subtracted) from either variable, a new line must be drawn to fit the points, but its slope (its angle of inclination, representing the change in Y for each unit change in X) is the same.<sup>42</sup>

2) When there is only random error in the dependent variable, the impact (measured by its slope) of the independent variable is unchanged, but we become less confident that the impact is really there. Part d of figure 2.3 illustrates this result: the same line fits both sets of points, but does not fit them equally well. If we find a significant relationship in spite of suspected random error in the dependent variable, it is actually *more* impressive than detecting a relationship when there is no such error. By the same reasoning, if we fail to find a relationship, we need not abandon hope, because the relationship may be hidden by the measurement error.

3) When there is only random error in the independent variable, the slope estimate *is* biased, as well as inefficient. But in all cases the slope is biased toward zero, as shown in part 3 of figure

2.3. Once again, if we find a significant effect in spite of suspected random error in the independent variable, it is all the more impressive; if we find no relationship, then it may be that random error is hiding it.

This analysis suggests that what seems like a big problem for descriptive inference is less of a problem for causal inference. To the extent that there is systematic error, it has no effect on our estimates of the marginal effect of any cause on democracy. To the extent that there is random error, it should increase confidence in the many findings that have achieved statistical significance in spite of the error, although it does cloud the interpretation of other findings that are only marginally or occasionally significant. It might seem, therefore, that if criticisms of measurement error in democracy indicators were intended to undermine confidence in analyses of the causes of democracy, then they have backfired: the worse the measurement, the more we should believe any significant findings, and the more we should give the benefit of the doubt to findings that are marginally insignificant.

This is true in principle. In practice, however, it is very unlikely that systematic measurement error is limited to adding or subtracting constants or that what we treat as random error is purely random. It is more likely that we overrate democracy in some clusters of countries and underrate it in others, and that there is a pattern to our over- and under-ratings, as Bollen found with respect to various regions. If so, this is neither purely random nor purely systematic measurement error, but error that is correlated with some unspecified explanatory factor. It is always tempting to dismiss as random variation whatever aspects of democracy that are not well accounted for by our explanations, but it is far more likely that we have not yet discovered the keys to those aspects that would reveal their systematic components. Rather than discuss these

issues under the heading of measurement error, however, I will save them for later examination of model specification and omitted variable bias in chapter 7.

### *Levels of Measurement*

In addition to their consequences for case and model selection and inferences, choices about concepts influence the level of measurement that can be attempted. Thick concepts are most easily measured with nominal or ordinal data, while thin concepts lend themselves more naturally to interval or ratio data.

Ratio indicators are numerical measures with a natural zero point and intervals of equal size. Because the intervals are of uniform size, an increase from 2 to 3 is equivalent to an increase from 100 to 101. That is, a one-unit increase is the same no matter what the initial value is. Because ratio indicators have an absolute zero, ratios are also meaningful: a score of 4 is twice as large as a score of 2. Few democracy indicators are truly at the ratio level of measurement, but one good example is Bollen's suffrage indicator.<sup>43</sup> This variable measures the percentage of the adult population that is entitled to vote. An increase from 25 to 50 percent of the adult population is an increase of 25 percentage points, as is an increase from 50 to 75 percent. It also makes sense to say that a country with a score of 90 percent has suffrage that is three times as extensive as that of a country with a score of 30 percent. Another example is Vanhanen's Index of Democratization, which is the product of turnout and the proportion of the legislative vote not won by the largest party.<sup>44</sup> A zero on this index would mean that either no one voted in the election or that the largest party received all the votes. Many independent variables are ratio data, especially socioeconomic variables such as per capita Gross Domestic Product, economic growth rates, inflation rates, and literacy rates.<sup>45</sup>

Interval indicators, like ratio indicators, have equal intervals but, unlike ratio indicators, lack an absolute zero point. Outside of the political world, the best example is the Fahrenheit temperature scale. An increase of 10 degrees is an increase of 10 degrees no matter what the starting temperature is; but it is not correct to say that 90 degrees is three times as hot as 30 degrees. One interval-level democracy indicator is Bollen's Liberal Democracy Series. This variable takes on many possible values between 0 and 100, so its intervals can be assumed to be approximately equal.<sup>46</sup> However, a zero on this indicator does not qualify it as ratio data unless it corresponds to situations of zero liberal democracy, i.e., the complete absence of liberal democracy. Highly repressive regimes such as that of the Soviet Union under Stalin, Nazi Germany, China during the Cultural Revolution, Cambodia under the Khmer Rouge, or North Korea could be considered antitheses of liberal democracy, but it is hard to imagine no freedom of any kind, the complete exclusion of all citizens from any role in policymaking, and the complete lack of any information whatsoever about politics.

Ordinal indicators are rankings: they reflect relative positions on some dimension, but the distances between ranks are assumed to be unknown. The third-best score could be a close third or a distant third; the top score could be the best by far or nearly a tie with the second-best score. Most democracy indicators are technically at the ordinal level of measurement.<sup>47</sup> A good example is Mainwaring, Brinks, and Pérez Liñán's classification of Latin American regimes as "democratic," "semidemocratic," or "authoritarian": three ranges on an underlying dimension from democratic to authoritarian.<sup>48</sup> Others include Wesson's Democracy Classification (5 ranks), the Freedom House Indexes of Political Rights and Civil Liberties (7 ranks each), Gurr's Polity Indexes of Democracy and Autocracy (10 ranks each), the Coppedge-Reinicke Polyarchy Scale

(11 ranks), and Hadenius's Index of Democracy. When an ordinal indicator has a small number of ranks, its ordinal nature should be respected. However, the more ranks it has, the more closely it approximates interval data.<sup>49</sup> For this reason, scholars have often combined the two Freedom House indicators into an index ranging from 2 (least freedom) to 14 (most freedom) and the two Polity variables into an index ranging from -10 (full autocracy) to +10 (full democracy). Hadenius's index has so many levels between 0 and 10 (such as 7.2, 9.4, etc.) that he considered it accurate to within 0.6 points.<sup>50</sup>

Classifications, labels, or typologies that do not imply any ranking on an underlying dimension are nominal indicators. For example, Hannan and Carroll categorized regimes as multi-party, one-party, military, or traditional no-party regimes. Similarly, Linz and Stepan defined a typology of democratic, authoritarian, totalitarian, post-totalitarian, and sultanistic regimes. In both examples, the first regime in each list is democratic, but there is no implicit ordering of the other regimes on a democracy-nondemocracy dimension.<sup>51</sup> Nominal indicators contain the least amount of quantitative information, but they often compensate by representing a great deal of qualitative information. Nominal indicators can even be used to measure multidimensional phenomena, simply by establishing thresholds on each dimension and labels or types that correspond to being above or below the threshold on certain dimensions. Figure 2.2 shows all the nominal classifications that might be generated with Linz's four criteria. Linz himself labeled three of these types that lie on a democracy-totalitarianism continuum, but if the other possible types were labeled, there would be no underlying ordering.

Dichotomies are the simplest nominal indicators, in which all cases are either above the threshold on all the criteria or assigned to a residual category. Any list of democracies qualifies as

a dichotomous nominal indicator. The best recent example is the classification of “democracies” and “dictatorships” by Alvarez, Cheibub, Limongi, and Przeworski.<sup>52</sup> If a regime had an elected executive and legislature, more than one party, and the opposition had a real chance to win the next election, then they coded it a “democracy”; otherwise, they assigned it to a residual (i.e., leftover) category of “dictatorships.” Note that although dichotomies can handle thick criteria, this particular indicator is based on rather thin criteria.

No level of measurement is inherently superior to another. Nevertheless, there have been heated debates about whether democracy should be measured by dichotomous or continuous indicators. On one side, Sartori and Przeworski argue that democracy is inherently dichotomous: a country is either democratic or it is not; there are no degrees of democracy.<sup>53</sup> Therefore, measuring democracy as a matter of degree implies conceptual confusion and increases measurement error. On the other side, Dahl, Bollen, and others argue that there are degrees of democracy and a continuum of “democraticness” ranging from very democratic to highly undemocratic regimes.<sup>54</sup> For them, higher levels of measurement improve the accuracy and reliability of democracy indicators. I agree more with Collier and Adcock, who argue that almost any concept can be thought of as either categorical or continuous.<sup>55</sup> Counter to the best-known example of a supposedly inherent dichotomy, it is not strictly true that a woman cannot be half pregnant, for it depends on how one defines “pregnant.” She can be 4.5 months pregnant; she can have delivered one of two twins; she can, for a brief moment during labor, have the baby half in and half out; she can be heading for a miscarriage or a stillbirth; and so on. If pregnancy can be a matter of degree, so can anything else. The real issue is not whether a concept is *a priori* categorical or continuous, but which level of measurement is most useful for the analysis one wishes to do.

### *Precision*

The usefulness of an indicator depends on how valid and reliable it is, which we have already discussed; but it also depends on how precise the indicator is. Precision is a criterion separate from considerations of validity, reliability, and level of measurement.<sup>56</sup> Precision is the fineness of the distinctions made by an indicator: the amount of detail. Measurements can be precise or imprecise whether they are quantitative or qualitative. A statement that a country is 91.8 percent democratic would be extremely precise in quantitative terms (if such a statement could be made reliably). But it would also be very qualitatively precise to describe that country's democratic institutions in sufficient detail to establish that the d'Hondt system of proportional representation is used in a single national district for legislative elections, opposition parties receive equal broadcast time during campaigns, citizens legislate directly in referendums several times a year, city council meetings are open to the public, and so on.

There is, in practice, a tradeoff between quantitative and qualitative precision. Quantitative precision usually entails a loss of qualitative information, and qualitative precision usually entails a loss of quantitative information. If both continuous and categorical indicators measured exactly the same concept, then we would prefer the continuous one on the grounds that it is more informative, more flexible, and better suited for sophisticated testing. For example, if the concept of interest were “breadth of the suffrage” we might choose between two indicators: a qualitative indicator that divided countries into two categories such as “universal adult suffrage” and “suffrage with restrictions”; or a quantitative index of the percentage of the adult population that is eligible to vote. Of these two, we should prefer the quantitative indicator because it measures the concept with finer gradations, which give us more quantitative information. If one wanted a

categorical measure, it could always be derived from the continuous one by identifying one or more thresholds that correspond to the categories desired, such as “at least 95 percent of adults are eligible to vote.” A dichotomized indicator would sort cases and interact with other variables the same way a dichotomy would--again, assuming that they measured exactly the same concept. The continuous indicator contains more information, which we could choose to ignore, but the reverse is not true: one cannot derive a continuous measure from a categorical one without adding new information about gradations.

However, this argument has a flip side: if a qualitative and a quantitative indicator measured a concept with equally fine gradations, we would prefer the qualitative indicator on the grounds that it provided more information about the qualities that are being represented. Let us suppose that we have, on the one hand, a three-fold typology dividing regimes into democratic, authoritarian, and totalitarian regimes; and on the other hand, a three-point scale of, say, “degrees of accountability.” In this example, we could derive a quantitative indicator from the qualitative typology, but we could not derive the typology from the accountability indicator without adding qualitative information about regime qualities beyond “accountability.”

Quantitative precision affects how appropriate an indicator is for the kind of analysis one intends to carry out.<sup>57</sup> There is a hierarchy among the levels of measurement based on the kinds of mathematical operations that can be meaningfully performed with them. Nominal measurements can only be used for identity relations; ordinal measurements can establish identity and inequalities; interval measurements are useful for identities and inequalities and can be counted or added and subtracted; and ratio measurements are useful for identities and inequalities and can be counted, added and subtracted, multiplied and divided, and subjected to higher-order

transformations such as logarithms and exponentials. These possibilities for mathematical manipulation constrain the ways indicators can be used in quantitative analysis. Table 2.6 lists various types of descriptive and explanatory quantitative analysis that are appropriate for each level of measurement.

These analytic constraints have important implications for democratization research. On the one hand, dichotomies--because they are categorical and because they can take into account multiple criteria--correspond most naturally to the concept of "regimes," which can persist without relevant alteration for years. Dichotomies therefore lend themselves to analyses of rare and dramatic changes such as democratic transitions and breakdowns and to the related concept of regime "life expectancy," as in the important work by Przeworski et al. On the other hand, analysis of subtle, short-term, or partial changes in democracy such as political liberalization, democratic deepening, institutional crisis, and quality of democracy require a higher level of measurement.

### **Thickening Thin Concepts**

It is tempting to conclude that different types of measurement are appropriate for different kinds of research and that there is no "best" kind of measurement. And again, the problem with that view is that it impedes the cumulation of knowledge. Qualitative and quantitative researchers have no choice but to talk past each other as long as their evidence measures qualitatively different concepts. Therefore, there is a great need to overcome this division. It can be done by developing quantitative indicators of thick concepts.

The idea may be offensive to those who are comfortable with fine qualitative distinctions and distrust numbers. Their attitude is reminiscent of skeptics who argued years ago that one

could not reduce, for example, Beethoven to a string of numbers. Now it can be done, and is done, on compact disks. With enough technology, laboriously developed over a century at great expense, we can sample multiple frequencies thousands of times per second, convert it into digital code, and then reproduce the sound so well that it is virtually indistinguishable from "Beethoven."

In social science, we already do something like this with dichotomies. Any dichotomous concept can be perfectly operationalized as a dummy variable, which takes on values of zero or one. We can pile as many components as we like onto a dummy variable and still represent them with these two values without suffering any loss of information. The components do not even have to be unidimensional, because one cutpoint can be picked on each component and the dummy defined to equal "1" only when every component equals "1." This is the exact mathematical equivalent of a multifaceted, categorical distinction. Quantitative indicators do not strip away qualitative meaning; rather, they establish a *correspondence* between meaningful qualitative information and numbers.

In principle we should also be able to create polychotomous, ordinal, interval, or (in some cases) ratio-data indicators of thick concepts. The challenge is threefold. The first challenge is to ensure that every element that contributes to the definition of a thick concept is measured by a quantitative variable. The second challenge is to reconceptualize each of these elements as a matter of degree, not as just as an either/or difference.

The third challenge in bringing about the best of the qualitative and quantitative approaches is to preserve the structure of the qualitative concept. This requires grouping components into dimensions correctly and combining them into a single index for each dimension. First the analyst breaks the "mother" concept up into as many simple and relatively

objective components as possible. Second, each of these components is measured separately. Third, the analyst examines the strength of association among the components to discover how many dimensions are represented among them and in the mother concept. Fourth, components that are very strongly associated with one another are treated as unidimensional, i.e., as all measuring the same underlying dimension, and may be combined. Any other components or clusters of components are treated as indicators of different dimensions. If the mother concept turns out to be multidimensional, the analyst then has two or more unidimensional indicators that together can capture its complexity. If the mother concept turns out to be unidimensional, then the analyst has several closely associated component indicators that may be combined into a single indicator that captures all the aspects of that dimension better than any one component would.<sup>58</sup>

I suspect that we are not likely to achieve much improvement in reliable and valid measurement until we begin working with a thicker, multidimensional concept of democracy. If democracy is multidimensional, then democracy indicators must be multidimensional as well; otherwise, measurements are compromised by measurement error or validity problems. The worst tactic for coping with multidimensionality is to assume blindly that all the components are unidimensional and barrel on, adding or averaging these apples and oranges. The fruit of such efforts may turn out to be reasonable at the extremes, but is likely to be a meaningless mess in the middle.

A more acceptable tactic is to tolerate a low level of measurement: interval rather than ratio data, ordinal rather than interval, a 3-point scale rather than a 10-point scale, or a dichotomy rather than a scale. This tactic is available because unidimensionality is a matter of degree.

Sometimes dimensions are distinct but parallel, or “bundled.” The tighter the bundle, the less measurement error is created when they are combined simply into an allegedly unidimensional indicator. If one is content to produce an indicator of democracy at a low level of measurement—say, a 3-point scale of democracies, semidemocracies, and non-democracies—one can aggregate components that lie on different and fairly weakly correlated dimensions.

As noted above, dichotomies are the limiting case of this tactic. But dichotomizing is radical surgery. It amputates every dimension below the cutoff and tosses all that information into a residual bin labeled “non-democracy.” If this information is truly not worth knowing, such radical surgery can be justified—for example, if it is the only way to salvage a viable indicator. But if there is serious doubt about where to cut, caution is advised.

Obviously, we are far from creating all the rich data that would be needed to measure any thick concept of democracy in a large sample. Comparative politics is scandalously data-poor, and the problem is not limited to democratization research. Correcting the situation would take an enormous investment in rigorous, systematic data collection on a large scale. Resources to make it possible may not be available now, but in order to obtain the resources it is first necessary to decide that such data are meaningful, desirable, and, in principle, feasible to create. In the meantime, it is useful to keep in mind even today that small- and large-N analysis, thick and thin, are parts of a whole, and that as data collection improves, we can expect them to converge rather than diverge into entirely separate camps.

### **Conclusion**

Democracy can be measured, and has been measured, in many different ways. However, the indicators in our possession today capture only a thin version of democracy. Despite the fact

that democracy is demonstrably a multidimensional phenomenon, and probably more multidimensional the more richly it is defined, most existing indicators focus on just one of its dimensions--contestation. The bright side is that contestation has been measured fairly well, for almost all countries, over long and growing spans of time. As far as we know, the best existing indicators are not grossly biased. They may not be sufficiently reliable to be useful for intra-regional comparisons, but to the extent that there is measurement error, it does not seem to pose much of a problem for research on causes of democracy in large and diverse samples. In fact, the knowledge that there is some measurement error should actually increase our confidence in findings that turn out to be statistically significant in spite of such error. Nevertheless, we need thicker indicators, over a longer span of time, with greater attention to reliability and additional dimensions of democracy.

**Table 2.1: Types and Subtypes of Democracy**

- A. Economic Democracy
- B. Social Democracy
- C. Communitarian Democracy
- D. Political Democracy
  - 1. Procedural Democracy
    - a. Direct Participatory Democracy
    - b. Representative Democracy
      - 1) Popular Sovereignty Democracy
      - 2) Liberal Democracy

**Table 2.2: Elements of Held's Models of Democracy****INSTITUTIONS**

regular elections  
 elections for many offices  
 secret ballot  
 strong executive  
 party politics  
 one person, one vote  
 multiple or different voting rights  
 representation  
 constitutional limits to state power  
 separation of powers/checks and balances  
 rule of law  
 internal party democracy  
 mixed government  
 direct participation in decision making  
 some appointments by lot  
 strict term limits  
 payment for participation  
 public campaign finance  
 innovative feedback mechanisms  
 universal adult suffrage  
 proportional representation  
 independent, professional bureaucracy  
 professional bureaucracy  
 limited bureaucracy  
 guarantees of civil liberties  
 guarantees of political rights  
 workplace democracy  
 minimization of unaccountable power centers  
 representation of corporate interests  
 representation of the powerful  
 restriction of some interest groups  
 jury service

**INTERNATIONAL SYSTEM**

global state  
 international competition  
 pluralist, free-market international order  
 unequal international order

**SOCIAL STRUCTURE**

small community  
 patriarchal family or society  
 intense societal conflict  
 autonomous civil society  
 free-market society  
 maintenance of religious worship  
 interest-group pluralism

**ECONOMIC SYSTEM**

private property  
 market economy  
 industrial society  
 non-industrial society  
 economic inequality  
 priority of economic interests  
 exclusion of some from effective participation  
 by economic inequalities  
 redistribution of resources  
 experiments with collective property

**CULTURE AND PARTICIPATION**

public debates  
 participation in local government  
 competition for power  
 openness to institutional reform  
 transparency  
 no distinction between citizens and officials  
 individualism  
 poorly informed or emotional voters  
 culture of toleration  
 consensus on legitimate scope of politics  
 procedural consensus  
 moderate level of participation  
 liberal leadership

**MISCELLANEOUS**

strong leadership  
 popular sovereignty  
 unbiased state  
 state with interests of its own  
 large nation-state  
 right to childcare  
 demilitarization

Source: Author's compilation of elements discussed in David Held, *Models of Democracy*, 2<sup>nd</sup> ed. (Stanford: Stanford University Press, 1996).

**Table 2.3: Definitions of Authoritarian Regime and a Low Degree of Polyarchy Contrasted**

Authoritarian Regime:	Polyarchy Scale Score 5:
[Political systems without] free competition between leaders to validate at regular intervals by nonviolent means their claim to rule. . . <sup>59</sup>	[There are] no meaningful elections: elections without choice of candidates or parties, or no elections at all.
. . . political systems with limited, not responsible, political pluralism	Some political parties are banned and trade unions or interest groups are harassed or banned, but membership in some alternatives to official organizations is permitted. Dissent is discouraged, whether by informal pressure or by systematic censorship, but control is incomplete. The extent of control may range from selective punishment of dissidents on a limited number of issues to a situation in which only determined critics manage to make themselves heard. There is some freedom of private discussion. Alternative sources of information are widely available but government versions are presented in preferential fashion. this may be the result of partiality in and greater availability of government-controlled media; selective closure, punishment, harassment, or censorship of dissident reporters, publishers, or broadcasters; or mild self-censorship resulting from any of these.
without elaborate and guiding ideology, but with distinctive mentalities	
without extensive nor intensive political mobilization, except at some points in their development	
and in which a leader or occasionally a small group exercises power within formally ill-defined limits but actually quite predictable ones.	
Source: Juan J. Linz, "Nondemocratic Regimes," in Fred I. Greenstein and Nelson W. Polsby, eds., <i>Handbook of Political Science</i> , v. 3: <i>Macropolitical Theory</i> , (Reading, Mass.: Addison-Wesley, 1975), p. 264.	Source: Michael Coppedge and Wolfgang Reinicke, "Measuring Polyarchy," <i>Studies in Comparative International Development</i> 25:1 (Spring 1990): 53-54.

**Table 2.4: Two Unidimensional Components of Polyarchy**

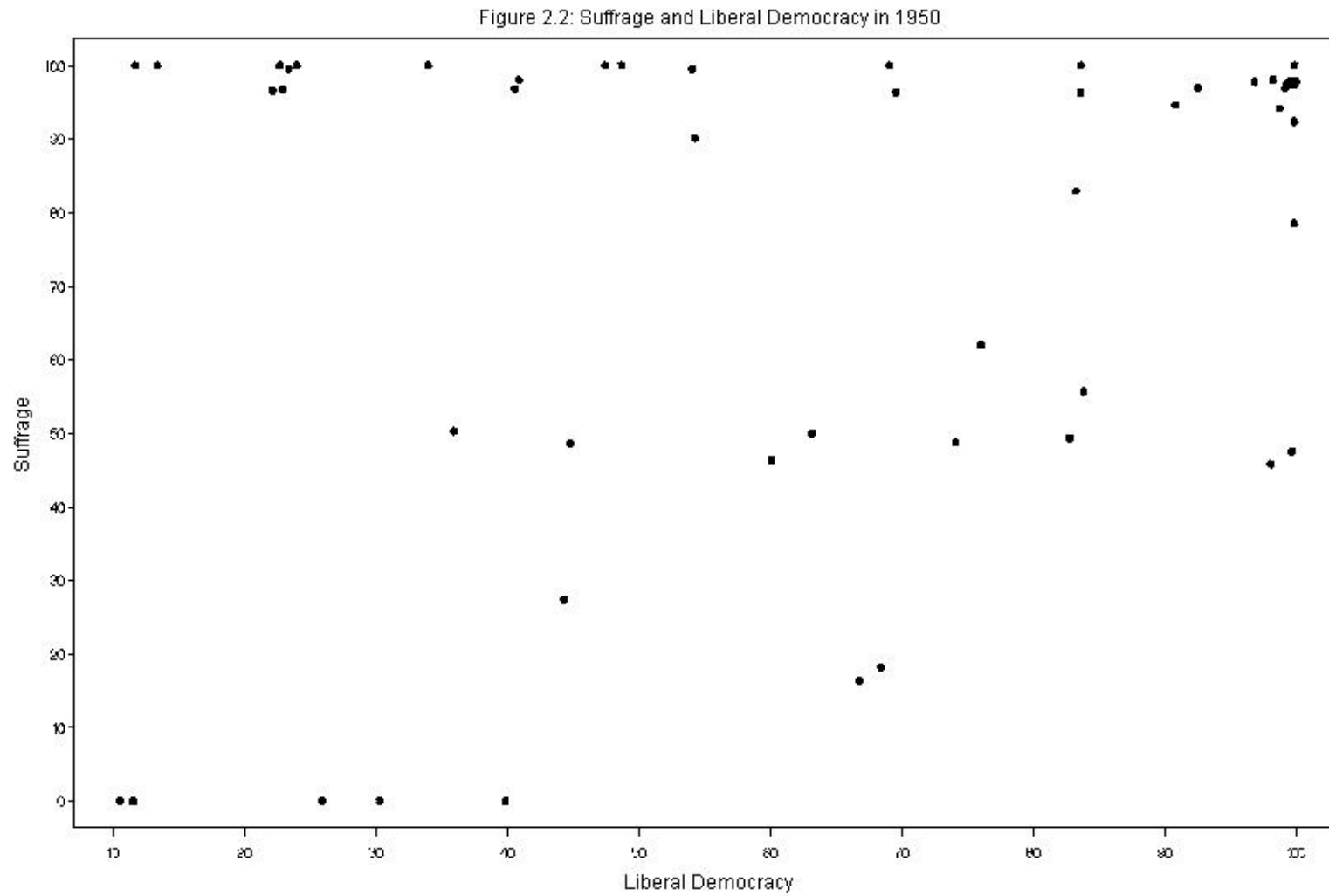
Numbers in cells are the number of countries that had both the row and the column characteristic in 1985.

		Freedom of Expression		
		free expression	dissent discouraged	dissent forbidden
Media Pluralism	media are diverse and protected by law	40	3	0
	media have a pro-government bias	11	24	0
	diversity allowed only when harmless	0	45	1
	complete official domination	0	6	29

Source: Data used in Michael Coppedge and Wolfgang Reinicke, "Measuring Polyarchy," *Studies in Comparative International Development* 25:1 (Spring 1990): 51-72, and available on line at <http://www.nd.edu/~mcoppedg/crd>.

**Figure 2.1: Intension and Extension of Linz's Definitions of Regime Types**

ideology of leaders	participa-tion	institution-alized?	full pluralism		limited pluralism		monistic control	
			elections	not	elections	not	elections	not
indetermin-ate ideology	welcome but not forced	yes	Democratic					
		no						
	discouraged	yes						
		no						
	forced	yes						
		no						
distinctive mentality	welcome but not forced	yes						
		no						
	discouraged	yes				Authoritarian		
		no						
	forced	yes						
		no						
elaborate and guiding ideology	welcome but not forced	yes						
		no						
	discouraged	yes						
		no						
	forced	yes					Totalitarian	
		no						



Source: Kenneth A. Bollen, "Cross-National Indicators of Political Democracy, 1950-1990" [computer file]. 2nd ICPSR version. Chapel Hill, NC [producer], 1998. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2001.

**Table 2.5: Average Correlations among Democracy Indicators for 1985**

Indicator	Mean Correlation with the other indicators	Min	Max	Std. Dev.
Wesson's Democracy Classification	0.895	0.840	0.945	0.031
Bollen's Liberal Democracy Series	0.890	0.814	0.964	0.041
Freedom House Political Rights Index	0.886	0.832	0.945	0.039
Coppedge-Reinicke Polyarchy Scale	0.882	0.796	0.941	0.038
Coppedge and Reinicke's Multi-Party Elections Variable	0.873	0.792	0.939	0.046
Freedom House Civil Liberties Index	0.871	0.784	0.937	0.044
Gurr's Institutionalized Democracy Scale	0.865	0.760	0.942	0.050
Sussman's Print Media Control Variable	0.863	0.772	0.921	0.043
Humana's Freedom of Political Opposition Variable	0.862	0.736	0.934	0.047
Banks's Party Legitimacy Variable	0.860	0.786	0.964	0.045
Gurr's Competitiveness of Executive Recruitment Variable	0.851	0.764	0.942	0.048
Gurr's Competitiveness of Political Participation Variable	0.837	0.758	0.910	0.040
Vanhanen's Index of Democracy	0.827	0.759	0.878	0.033
Humana's Right of Assembly and Association Variable	0.826	0.701	0.883	0.046
Sussman's Broadcast Media Control Variable	0.823	0.729	0.889	0.037
Humana's Freedom of Information Variable	0.813	0.691	0.879	0.054
Gurr's Executive Constraints Variable	0.793	0.691	0.931	0.063
Number of countries included	137	89	175	31.4

Source: calculated from data in Kenneth A. Bollen, "Cross-National Indicators of Political Democracy, 1950-1990" [computer file]. 2<sup>nd</sup> ICPSR version. Chapel Hill, NC: University of North Carolina [producer], 1998. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001.

[Insert Figure 2.3 about here.]

**Table 2.6: Appropriate Uses of Indicators at Different Levels of Measurement**

	Dichotomous	Nominal	Ordinal	Interval	Ratio
<b>Description:</b>					
Percentage change					✓
Standard deviation and variance				✓	✓
Factor analysis			✓*	✓	✓
Mean			✓*	✓	✓
Median			✓		
Mode	✓	✓			
<b>As an independent variable:</b>					
Complex transformations					✓
Multiplicative interactions					✓
Differenced variable				✓	✓
Random variable				✓	✓
Dummy-variable interaction	✓	✓	✓		
Dummy variable	✓	✓	✓		
<b>As a dependent variable:</b>					
Regression				✓	✓
Product-moment correlation				✓	✓
Rank-order correlation			✓		
Ordered logit or probit			✓		
Multinomial logit		✓			
Cross-tabulation	✓	✓	✓		
Discriminant and log-linear analysis	✓				
Event-history and Boolean analysis	✓				

\*Technically, one should not calculate means or perform factor analysis with ordinal data. Nevertheless, we have become accustomed to the unnatural concept of an "average rank." Also, a factor analysis will work only to the degree that any ordinal variables in the analysis approximate interval data. In practice, they often do.

### Notes

1. W. B. Gallie, "Essentially Contested Concepts," *Proceedings of the Aristotelian Society* 56 (1956): 169.
2. See David Collier and Steven Levitsky, "Democracy with Adjectives: Conceptual Innovation in Comparative Research," *World Politics* 49:3 (April 1997): 430-51; Richard S. Katz, *Democracy and Elections* (Oxford, 1997), pp. 26-99; and my summary of David Held, *Models of Democracy*, 2<sup>nd</sup> ed. (Stanford: Stanford University Press, 1996), in Table 2.2, below. For quantitative indicators of democracy, see Kenneth A. Bollen, "Cross-National Indicators of Political Democracy, 1950-1990" [computer file]. 2nd ICPSR version. Chapel Hill, NC: University of North Carolina [producer], 1998. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2001. This source contains Bollen's nearly exhaustive compilation of 87 quantitative democracy indicators covering any of the years between 1950 and 1990.
3. This element is implied by Linz's explicit statements that authoritarian regimes are by definition nondemocratic. The language comes from his own definition of a democratic political system in Juan J. Linz, "Totalitarian and Authoritarian Regimes," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science*, v. 3: *Macropolitical Theory*, pp. 175-411 (Reading, Mass.: Addison-Wesley, 1975), pp. 182-3.
4. Linz, "Totalitarian and Authoritarian Regimes," p. 264, quoting the definition from his own "An Authoritarian Regime: The Case of Spain" from 1964.
5. Seymour Martin Lipset, "Some Social Requisites of Democracy," *American Political Science Review* 53 (1959): 69-105; Phillips Cutright, "National Political Development: Measurement and Analysis," *American Sociological Review* 28 (1963): 253-264; Phillips Cutright and James A.

- Wiley, "Modernization and Political Representation: 1927-1966," *Studies in Comparative International Development* 5 (1969): 23-44; Zehra F. Arat, "Democracy and Economic Development: Modernization Theory Revisited," *Comparative Politics* 21 (1988): 21-36.
6. Giovanni Sartori, "Concept Misformation in Comparative Politics," *American Political Science Review* 64:4 (December 1970): 1033-1053. See David Collier and Steven Levitsky, "Democracy with Adjectives: Conceptual Innovation in Comparative Research," *World Politics* 49 (April 1997): 430-51 for variations on this theme.
7. Linz wrote hundreds of pages describing political systems that differed from these basic three. Some, such as "authoritarian situations," could fit in Figure 2.2 without any revisions to the characteristics around which it is structured. Others, such as sultanistic regimes and post-totalitarian regimes, had defining characteristics that were not part of Figure 2.2 and therefore suggest that Linz's underlying classificatory scheme was still more complex.
8. Richard Katz, *Democracy and Elections* (New York: Oxford University Press, 1997), pp. 10-14.
9. Edward A. Purcell, Jr., *The Crisis of Democratic Theory: Scientific Naturalism and the Problem of Value* (U Press of Kentucky, 1973).
10. Robert A. Dahl, *Polyarchy* (New Haven: Yale University Press, 1970), p. 3. In his later work, *Democracy and Its Critics* (New Haven: Yale University Press, 1989), Dahl amended "government policies depend on votes and other expressions of preference" to "control over governmental decisions about policy is constitutionally vested in elected officials" ( p. 233). This small change in wording made more explicit Dahl's acceptance of representative democracy.
11. Freedom House claims to consider a very long list of characteristics in its ratings, but its

procedures are not transparent and its ratings are similar to other indicators that take few characteristics into account. Complete Freedom House ratings are available at <http://www.freedomhouse.org>.

12. Philippe C. Schmitter, and Terry Lynn Karl, "What Democracy Is . . . and Is Not," *Journal of Democracy* 2 (Summer 1991): 76-80.

13. Jonathan Hartlyn and Arturo Valenzuela, "Democracy in Latin America Since 1930," in Leslie Bethell, ed., *The Cambridge History of Latin America*, v. VI: *Latin America Since 1930: Economy, Society, and Politics* (Cambridge: Cambridge University Press, 1994); O'Donnell, "Delegative Democracy"; Larry Diamond, *Developing Democracy: Toward Consolidation* (Baltimore: Johns Hopkins University Press, 1996), pp. 111-112.

14. J. Samuel Valenzuela, "Democratic Consolidation in Post-Transitional Settings: Notion, Process, and Facilitating Conditions," in Scott Mainwaring, Guillermo O'Donnell, and J. Samuel Valenzuela, eds., *Issues in Democratic Consolidation: The New South American Democracies in Comparative Perspective* (Notre Dame, Ind.: University of Notre Dame Press, 1992), pp. 62-68; Guillermo O'Donnell, "On the State, Democratization, and Some Conceptual Problems: A Latin American View with Glances at Some Post-Communist Countries," *World Development* 21 (1993): 1355-69.

15. Guillermo O'Donnell, "Human Development, Human Rights and Democracy," paper prepared for the workshop on the "Quality of Democracy," sponsored by the United Nations Development Program, Regional Division for Latin America and the Caribbean (UNDP-DRALC), and the Proyecto Estado de la Nación, Costa Rica (2001).

16. Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in

John Worrall and Gregory Currie, eds., *The Methodology of Scientific Research Programmes*, pp. 8-101 (Cambridge: Cambridge University Press, 1978).

17. Fernando Henrique Cardoso and Enzo Faletto, *Dependencia y desarrollo en América Latina* (México: Siglo Veintiuno, 1971).

18. Guillermo O'Donnell, *Modernization and Bureaucratic-Authoritarianism : Studies in South American Politics* (Berkeley: Institute of International Studies, University of California, 1973).

8. Guillermo O'Donnell, "Delegative Democracy," *Journal of Democracy* 5 (April 1994): 57-64.

20. Guillermo O'Donnell, "Horizontal Accountability and New Polyarchies," Kellogg Institute Working Paper No. 253 (April 1998).

21. Ruth Berins Collier and David Collier, *Shaping the Political Arena* (Princeton: Princeton University Press, 1991).

22. Andrew C. Gould, "Conflicting Imperatives and Concept Formation," *Review of Politics* 61:3 (Summer 1999): 439-463.

23. Because thick concepts contain more ambitious theory, they should be subjected to testing, just as theories are. Calling a theory a concept does not render it immune to testing. Thin concepts, in contrast, are less theoretically ambitious; they assume less (and say less), and therefore leave more to induction. The thinner the concept, the less testing is required to achieve a similar level of readiness for theory-building.

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

25. Dahl, *Polyarchy*, p. 4.

26. After 1950, most countries eliminated suffrage restrictions, so this two-dimensional pattern is

less obvious for later years. Nevertheless, it is still two-dimensional in the sense that there is no empirical relationship between suffrage and contestation: most countries have full suffrage even if contestation is so low that elections are not fair or meaningful.

27. Arend Lijphart, *Patterns of Democracy* (New Haven, Conn.: Yale University Press, 1999), pp. 243-50.

28. For an even thicker definition of the quality of democracy that includes the satisfaction of basic human needs and respectful treatment of citizens by fellow citizens and the state, see Guillermo O'Donnell, "Human Development, Human Rights and Democracy," Working Paper No. 1, Workshop on "Calidad de la democracia y desarrollo humano en América Latina," Hotel La Condesa, Heredia, Costa Rica, February 1-2, 2002; and Proyecto Estado de la Nación, *Informe de la auditoría ciudadana sobre la calidad de la democracia*, vol. 1 (San José, Costa Rica: Proyecto Estado de la Nación, 2001).

29. Michael Coppedge and Wolfgang Reinicke, "A Scale of Polyarchy," *Studies in Comparative International Development* 25:1 (Spring): 51-72.

30. Bollen, "Cross-National Indicators of Political Democracy"; Gerardo Munck, "Measurement: Selecting Indicators and Generating Data," in *Report on Democratic Development in Latin America* (New York: Regional Bureau for Latin America and the Caribbean of the United Nations Development Program (UNDP-RBLAC), Inter-American Development Bank, and International IDEA, forthcoming 2003); Thorsten Beck, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh, "New Tools and new Tests in Comparative Political Economy: The Database on Political Institutions," at [http://www.worldbank.org/research/growth/political\\_data1.htm](http://www.worldbank.org/research/growth/political_data1.htm); Transparency International Corruption Perceptions Index, available at

<http://www.transparency.org/cpi/index.html>.

31. Gerardo L. Munck and Jay Verkuilen, "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices," *Comparative Political Studies* 35:1 (February 2002): 5-34. As Munck and Verkuilen note, such a theory should also be a tested and confirmed theory.

32. S.S. Stevens, "On the Theory of Scales of Measurement," *Science* 103 (1946): 677-680. Differences on interval indicators become ratio data, so for some applications interval-level measurement is a legitimate starting point for higher-level mathematical operations.

33. Bollen's index first calculates an indicator that we can call "political and legislative rights," composed of an average of a Freedom House-based Political Rights variable and the product of Legislative Effectiveness and Legislative Selection, two variables from the Arthur Banks dataset. All are weighted so that this average ranges from 0 to 100. If this indicator of political and legislative rights is less than the percentage of adults enjoying suffrage, then Liberal Democracy is the average of Political and Legislative Rights, on the one hand, and Banks's Party Legitimacy variable. However, if Political and Legislative Rights is greater than Suffrage, then Liberal Democracy is simply the average of Suffrage and Party Legitimacy. If Suffrage is high, therefore, Liberal Democracy is a weighted average of several contestation variables; and if Suffrage is very low, Suffrage drags down the Liberal Democracy series more powerfully. Bollen, "Cross-National Indicators of Political Democracy," p. 26.

34. This theory would imply a formula something like  $[kF+(1-k)P]S$ , where F is an indicator of individual freedom, P is an indicator of competition to make public policy, S is the probability that the state will faithfully execute policies, and k is the maximum possible height of the floor, that is, the highest degree of democracy a regime can provide if it does not allow elections. All

variables are scaled to a 0-1 interval, although  $k$  should be small, around 0.25.

35. A 1981 event-history analysis of 90 countries from 1950 to 1975 was ahead of its time, and the time series on which it was based has never been used in other research. See Michael T. Hannan and Glenn R. Carroll, "Dynamics of Formal Political Structure: An Event-History Analysis," *American Sociological Review* 46 (1981): 19-35. No other time-series analysis was published for another seven years. The next one was Lev S. Gonick and Robert M. Rosh, "The Structural Constraints of the World-Economy on National Political Development," *Comparative Political Studies* 21 (1988): 171-99.

36. Przeworski et al., *Democracy and Development*, p. 28.

37. Jonathan Hartlyn, "Contemporary Latin America, Democracy, and Consolidation: Unexpected Patterns, Re-elaborated Concepts, Multiple Components," in Joseph Tulchin and Amelia Brown, eds., *Democratic Governance and Social Inequality* (Boulder, CO: Lynne Rienner Publishers, 2002), pp. 103-30.

38. Kenneth A. Bollen, and Pamela Paxton, "Detection and Determinants of Bias in Subjective Measures," *American Sociological Review* 63 (June 1998): 465-78. This article replicates and extends somewhat Kenneth Bollen, "Liberal Democracy: Validity and Method Factors in Cross-National Measures," *American Journal of Political Science* 37:4 (November 1993): 1207-30.

These articles report confirmatory factor analyses to estimate the source bias and random measurement error in eight indicators that were presumed to measure political liberties and democratic rule.

39. Bollen calculated validity as 100 percent minus the sum of the systematic and random measurement error percentages. The worst of the eight indicators were Banks's Chief Executive

Elected, with 76 percent random error, and Banks's Competitiveness of the Nomination Process, with 38.5 percent systematic error. At the most valid extreme, Sussman's Freedom of Print Media indicator had only 9 percent random error and Freedom House's Political Rights had 6.4 percent systematic error and no random error. It is important to keep in mind that these estimates have meaning only with respect to the indicators on which they are based.

40. Munck and Verkuilen ["Conceptualizing and Measuring Democracy"] have argued that high correlations should not be very reassuring for two reasons. First, one can obtain high correlations when two indicators agree at the extremes but are completely unrelated in the middle. This point reinforces my conclusion that these indicators are useful for large-sample comparisons, but not to measure small differences or changes. Second, correlations (or Bollen's confirmatory factor analysis, for that matter) cannot detect any systematic bias that is shared by all the indicators. I believe that there is such a systematic bias, but that it is the tendency to focus on contestation rather than other aspects of democracy; hence my caveats about interpreting these indicators.

41. Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994), p. 156.

42. If either variable is multiplied by a constant, the new estimate of the slope will be changed. To be precise, the estimated slope will be the true slope times the constant if the error is in the dependent variable, or the true slope divided by the constant if the error is in the independent variable. In practice, the only instances in which a variable is multiplied or divided by a constant occur when one changes the units of measurement, for example from miles to kilometers or from 1970 dollars to 1990 yen. In any case, the substantive interpretation of the slope would be unchanged, so these are trivial possibilities.

43. Bollen, "Cross-National Indicators of Political Democracy," p. 29.
44. Tatu Vanhanen, *The Process of Democratization* (New York: Crane-Russak, 1990), pp. 27-29. The validity of this index is dubious, but it is clearly ratio data.
45. I have suggested above that democracy could be reconceptualized in terms of probabilities. If so, one could imagine a zero probability of free citizens influencing policy decisions. We lack such an indicator at present, but if one were created, it would qualify as ratio data.
46. This variable has more than 20 values, and Bollen made extraordinary efforts to make the sizes of differences in scores correspond to the sizes of differences in democraticness.
47. Gurr's and Hadenius's indicators are built from both interval- and ordinal-level components, but when the components are combined, the lower level of measurement prevails in the aggregated indicator.
48. Scott Mainwaring, Daniel Brinks, and Aníbal Pérez Liñán, "Classifying Political Regimes in Latin America, 1945-1999," Kellogg Institute Working Paper No. 280 (September 2000).
49. Sanford Labovitz, "The Assignment of Numbers to Rank Order Categories," *American Sociological Review* 35:3 (June 1970): 515-524. Scholars disagree about the number of ranks required for an ordinal indicator to approximate interval data. Some are comfortable with as few as 5; others prefer at least 20.
50. Axel Hadenius, *Democracy and Development* (Cambridge University Press, 1992), p. 76.
51. Hannan and Carroll, "Dynamics of Formal Political Structure," pp. 19-35; Juan J. Linz and Alfred Stepan, *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe* (Baltimore: Johns Hopkins, 1996), pp. 38-54.
52. Adam Przeworski, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi, *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*

(Cambridge University Press, 2000), pp. 13-36.

53. Giovanni Sartori, *The Theory of Democracy Revisited* (Chatham, NJ: Chatham House, 1987), p. 184; Michael Alvarez, José Antonio Cheibub, Fernando Limongi, and Adam Przeworski,

"Classifying Political Regimes," *Studies in Comparative International Development* 31 (2): 1-37.

54. Dahl, *Polyarchy*, pp. 231-235; Kenneth A. Bollen, "Political Democracy: Conceptual and Measurement Traps," *Studies in Comparative International Development* 25 (1990): 7-24.

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

56. One meaning of "precision" is "level of measurement." This is what Shively calls "precision in measurement." Because this definition can be confusing, what I mean by "precision" is what Shively calls "precision in *measures*." W. Phillips Shively, *The Craft of Political Research*, 4th ed. (Upper Saddle River, NJ: Prentice-Hall, 1998), pp. 53-70.

57. There is some debate about this. Some argue that numbers are numbers regardless of the measurement procedures that produced them, so any numbers can be used in any quantitative analysis. Others argue that the measurement theory that guided the generation of the numbers dictates the kinds of analysis that are appropriate. I lean toward the intermediate position that indicators can be used as though they were at a higher level of measurement when the measurement theory allows permits a reasonable interpretation of the results. This is not always possible, so scholars must be cautious. As Winkler and Hays observed, “. . . the road from objects to numbers may be easy, but the return trip from numbers to properties of objects is not.” Robert L. Winkler and William L. Hays, *Statistics: Probability, Inference, and Decision*, second edition (New York: Holt, Rinehart, and Winston, 1975), p. 282.

58. It is sometimes possible to combine multidimensional components into a single indicator.

Doing so, however, requires a theory that tells one how to combine them properly. In geometry, for example, “volume” is a single indicator of a multidimensional quality, but it cannot be calculated unless one knows the appropriate formula for the shape of the object in question.

59. This element is implied by Linz’s explicit statements that authoritarian regimes are by definition nondemocratic. The language comes from his own definition of a democratic political system (Linz 1975, 182-3).