

## **Thickening Thin Concepts: Issues in Large-N Data Generation**

This chapter clarifies the nature of the tradeoffs that are conventionally associated with two approaches to data generation in comparative politics. These two approaches, which I call "thick" and "thin," involve tradeoffs between generality and specificity, quantity and quality, and absolutes and matters of degree. I discuss the implications of these tradeoffs for conceptualization and measurement, using concepts and indicators of "democracy" as illustrations. Although the tradeoffs are familiar, there is no reason inherent in either approach for such tradeoffs to exist. Rather, the tradeoffs have been imposed on political scientists by practical limitations, especially the scarcity of appropriate data. The chasm between these approaches can and must be bridged, even though it would take an enormous effort to collect more data, different data, and better data than we currently have.

### **Conceptualization**

One fundamental difference between thick and thin approaches concerns their very building blocks—concepts. Thick concepts tend to be multifaceted, multidimensional, and imbued with theory; thin concepts tend to be simple, unidimensional, and more theoretically adaptable.

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 observed aspects. 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 1. I would not include all these items in even a thick definition of democracy, but at least one respected scholar considers them all relevant.

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>1</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>2</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.) 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.) 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 second difference in concepts concerns their dimensionality. 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. The higher the degree of association, the more reasonable it is to reduce the two components to one simple concept or a single dimension.

The Polyarchy Scale offers an example of such an empirical confirmation of unidimensionality.<sup>3</sup> It is composed of four components--indicators of fair elections, freedom of organization, freedom of expression, and pluralism in the media--which are all 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.

When a concept is multidimensional, its components do not vary together in this way. Intuitively, it is easy to imagine low-high or high-low combinations of 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. Dahl's concept of polyarchy again provides a good example, for contestation was only one of two dimensions in his concept. The other was participation (sometimes called inclusiveness), which he believed to vary independently of contestation. This supposition gave rise to his well-known diagram with closed hegemonies in one corner and polyarchies in the

opposite corner, but also mixed regimes called competitive oligarchies and inclusive hegemonies in the other two corners.<sup>4</sup>

The third difference concerns the relation of concepts to theory. 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>5</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>6</sup> Guillermo O'Donnell has formulated a series of others--bureaucratic-authoritarianism,<sup>7</sup> delegative democracy,<sup>8</sup> horizontal accountability.<sup>9</sup> The Colliers' "mode of incorporation" is yet another.<sup>10</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>11</sup> All of these

could be considered either very thick concepts or shorthand for theories.

One of the tradeoffs between thick and thin concepts was spelled out long ago by Sartori.<sup>12</sup> The more multifaceted a concept is (the broader its "intension"), the smaller the number of objects to which it applies (the narrower its "extension"). Although Sartori did not address the degree to which a concept is multidimensional and enmeshed in theory, these two qualities only reinforce his argument. The more baggage a concept must carry, the less widely it can travel. Kurt Weyland offers a fine example in this volume: "populism." If the term is equated only with a style of discourse exalting "the people," most Latin American politicians and many beyond the region would qualify as populists. But the more one adds on additional characteristics--spell-binding oratory from balconies, working-class support, neglect of party-building, redistributionist policies, military background, authoritarian proclivities--the fewer qualifying populists there are.

Thickness therefore adds meaning to a concept, but at the expense of wide applicability. Thin concepts have more general applicability, but tell us less about the objects they describe. Figure 1 illustrates this tradeoff 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. If these five characteristics are highly

unidimensional, this is not a problem, because most of the cases will fall in these three white cells. But if these regime types are multidimensional, then the regime types in some of the other cells must also be labeled.<sup>13</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.

This complementarity suggests that thick concepts are appropriate for small-N research, for which rich description is valued more than generalization, and thin concepts are appropriate for large-N work, in which generalization trumps detail. Unfortunately, such a rigid division of conceptual labor condemns scholars to talk past one another: when small-N and large-N analysts say "democratization," they mean different things. If theoretical knowledge is to accumulate, therefore, this conceptual chasm must be bridged. Doing so requires careful conceptual analysis: breaking down two concepts into their simplest common components to identify precisely how they overlap and how they differ. To do this, every element of a categorical definition can be reconceptualized as a threshold on a continuous dimension; these components can be measured separately, and then recombined to the extent that they are shown to be unidimensional. For example, if the Polyarchy Scale included all the components from Linz's definition of authoritarianism, then it would be a valid indicator of his concept, and it would have the

additional advantage of defining and measuring greater and lesser degrees of authoritarianism.<sup>14</sup>

No information would be lost, and some would be added.

There is another methodological implication in thick and thin concepts that receives less attention: as noted above, concepts also differ in how ready-made they are for theory-building. 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.

### **Thickening the Concept of Democracy**

Attempts to redefine "democracy" illustrate these tradeoffs. Because more and more developing countries now satisfy the rather minimalist existing requirements for democracy, it is difficult not to notice that some of these political systems have disturbing characteristics that seem intuitively inconsistent with 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>15</sup> Similarly, others stress the centrality of the rule of law and an independent judiciary.<sup>16</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>17</sup>

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 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.

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, for example. 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>18</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 redefine 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>19</sup> Such a five-dimensional concept could help us make meaningful distinctions among countries that satisfy the current minimal requirements for democracy. Of course, testing either the theories assumed by concepts or theories masquerading as concepts requires measurement.

### **Measurement**

Small-N research is conventionally associated with qualitative evidence and large-N with quantitative indicators. Small-N researchers are thought to be more concerned with differences in kind; they elaborately define types, which they combine into typologies. Large-N researchers are thought to be more concerned with differences of degree; they generate scores or values, which they combine into variables. However, this perceived distinction is an exaggerated stereotype. There are examples of large-N qualitative analysis, and small-N analysis can make use of quantitative data, as in voting behavior, survey research, and public policy or political economy.

But because the stereotype is common, a comparison of the merits of qualitative and quantitative evidence is germane.

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: “universal adult suffrage” or “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 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.

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.”

Once again, 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 is to get the structure of the concept right. The first challenge has already been discussed; the latter two require further explanation.

There are those who insist that some concepts are inherently categorical, and others inherently continuous. I agree more with Collier and Adcock (see chapter x), who argue that almost any concept can be thought of as either categorical or continuous. It is not strictly true, to counter the best-known example, 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.

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>20</sup>

### **Measuring Democracy**

Existing indicators of democracy are just beginning to satisfy rigorous standards for measurement. Democratic theorists before 1776 first simplified the task by progressively narrowing the concept, purging it of impractical components such as the appointment of administrators by lottery, and adapting it to the context of the large nation-state by accepting the idea of representation.<sup>21</sup> But from the French Revolution through Alexis de Tocqueville's *Democracy in America*, "democracy" was still so multifaceted that it was not even clearly distinct from social equality. The "elite theorists" during and after the Second World War then promoted an even narrower concept of democracy that was limited to political, rather than social or economic, components and did not require direct participation in policymaking, only in the selection of policymakers.<sup>22</sup> By the time political scientists began trying to measure democracy, the concept had therefore been reduced to selected national political institutions and practices and

some of their characteristics.

The first indicators of democracy had a few problems that required refinements. The early democracy indicators often confounded democracy with regime stability. In his classic 1959 article, for example, Lipset used the ordinal classifications “stable democracies/unstable democracies/ dictatorships” (for European and English-speaking countries) and “democracies/unstable dictatorships/stable dictatorships” (for Latin American countries).<sup>23</sup> Phillips Cutright’s index of “national political development” was the sum of a country’s democracy scores over a 21-year period, which made the number of years of democracy matter as much as the degree of democracy in each year.<sup>24</sup> As Kenneth Bollen has observed, this mistake has been repeated several times, even as late as 1988.<sup>25</sup> This is not to say that it is illegitimate to be interested in stable democracy. However, measuring stable democracy with anything more precise than an either-or category requires at least two dimensions, as regime stability and democracy vary independently: there are stable democracies, unstable democracies, stable nondemocracies, and unstable nondemocracies.

Other attempts to measure democracy excluded stability, sometimes by reporting a score for one time-point, sometimes by reporting an annual series of scores. But some of them compromised validity by including components that had little or no theoretical justification. For example, Vanhanen included the percentage of the vote won the governing party in his index of democracy, even though extremely fragmented party systems are not necessarily more democratic than two-party or moderate multiparty systems.<sup>26</sup> Another example is the Freedom House survey. Its checklists take into consideration the autonomy of elected representatives from military control, a country’s right of self-determination, citizens’ freedom from domination by economic

oligarchies, the autonomy of religious and ethnic minorities, gender equality, property rights, the freedom to choose family size, freedom from dependency on union leaders and bureaucrats, and freedom from gross government corruption, among other requirements.<sup>27</sup> Some of these components probably should not be included in a measure of democracy; others could be if the definition of democracy were fairly rich but should not be lumped together in the same index because they are likely to be multidimensional. Freedom House appears to combine its components in a flexible way that somehow avoids the worst potential biases, but it has not reported systematically how the components are related, so it is impossible for outside observers to confirm their validity or reliability.

Despite these measurement problems and another not yet mentioned, we know that even the relatively thin versions of democracy consist of at least two dimensions. For one of those dimensions we already have several indicators that are adequate for various large-N comparisons. One of Dahl's major contributions in *Polyarchy* was to argue convincingly that polyarchy has the two dimensions of contestation and inclusiveness.<sup>28</sup> He defined contestation as having several components, or institutional requirements--elected officials, free and fair elections, freedom of expression, associational autonomy, and the existence of alternative sources of information. Inclusiveness was defined solely in terms of the suffrage and widespread eligibility to run for public office. Coppedge and Reinicke later confirmed that the components of contestation are indeed unidimensional and may be legitimately combined into a single indicator, while the extent of the suffrage lies on a different dimension and should not be included as a component of contestation.<sup>29</sup> Many of the existing quantitative indicators of "democracy" are actually indicators of contestation. They are the Bollen Index of Political Democracy,<sup>30</sup> the Polity data on democracy



and autocracy,<sup>31</sup> the Freedom House ratings of Political Rights and Civil Liberties, the Polyarchy Scale, Hadenius' Index of Democracy,<sup>32</sup> and Bollen's Index of Liberal Democracy.<sup>33</sup> It has been demonstrated repeatedly that these indicators measure the same underlying dimension. Their intercorrelations, for example, usually exceed .83.<sup>34</sup> For an excellent survey and evaluation of the existing indicators, see the chapter in this volume by Munck and Verkuilen.

The indicators we have are by no means perfect. Bollen has demonstrated, for example, that Freedom House ratings, at least for 1979-1981, tended to underrate Eastern European countries and overrate Latin American countries by a small but statistically significant amount.<sup>35</sup> His index for 1980, which corrects for these biases as well as anyone can at this point, is probably the most valid indicator available today. But Bollen's index is a point measure; only a few are time-series, and unfortunately, as Munck and Verkuilen observe, the indicators with the longest historical coverage tend to be the ones with the most worrisome methodological flaws. Nevertheless, if one needs time-series data, there is little reason to avoid using—with reasonable cautions—the Alvarez-Cheibub-Limongi-Przeworski (ACLP), Freedom House or Polity data. If a dichotomous indicator is appropriate for one's purposes, the ACLP data are currently the best time-series. If a graded indicator is needed, despite violations of methodological canons during construction, the Freedom House and Polity data are good enough for large-scale comparative work involving a thin concept of democracy. According to Bollen's estimates, the Freedom House Political Rights ratings for 1979-1981 were 93 percent valid despite the regional bias. It also correlates at .938 with the Polyarchy Scale. These results suggest that we can expect very similar results from an analysis regardless of which of these indicators is used.<sup>36</sup> Of course, as Munck and Verkuilen point out, intercorrelations are not entirely reassuring because all indicators may well be biased in the same way, and because correlations may reflect agreement at the extremes more than agreement about the more difficult intermediate

cases.

However, 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. (See the chapters by Collier

and Adcock and Elkins.)

There are two strategies besides dichotomizing for coping with multidimensionality. The easier of the two is simply to develop a different indicator for each dimension 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 start measuring its component dimensions instead.

This disaggregated strategy would not amount to an admission of defeat. If we had separate indicators of different dimensions of democracy, we could explore empirically their interrelationships, 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 coping strategy: appropriate aggregation of components into a single indicator of democracy. We are not yet ready to do this for a multidimensional concept of democracy.<sup>37</sup> Doing so requires a stronger theory (and as Munck and Verkuilen note, a tested and confirmed 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. 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>38</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.

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.

**Table 1: Elements of Held's Models of Democracy**

regular elections	proportional representation
elections for many offices	independent, professional bureaucracy
secret ballot	unbiased state
public debates	state with interests of its own
jury service	individualism
strong executive	representation of corporate interests
strong leadership	private property
party politics	market economy
one person, one vote	patriarchal family or society
multiple or different voting rights	large nation-state
representation	global state
constitutional limits to state power	international competition
separation of powers/checks and balances	professional bureaucracy
participation in local government	experiments with collective property
guarantees of civil liberties	industrial society
guarantees of political rights	non-industrial society
competition for power	poorly informed or emotional voters
interest-group pluralism	culture of toleration
autonomous civil society	economic inequality
rule of law	procedural consensus
free-market society	consensus on legitimate scope of politics
workplace democracy	priority of economic interests
internal party democracy	moderate level of participation
openness to institutional reform	exclusion of some from effective
transparency	participation by economic inequalities
representation of the powerful	pluralist, free-market international order
mixed government	unequal international order
direct participation in decisionmaking	liberal leadership
some appointments by lot	limited bureaucracy
no distinction between citizens and officials	restriction of some interest groups
strict term limits	redistribution of resources
payment for participation	right to childcare
public campaign finance	maintenance of religious worship
innovative feedback mechanisms	intense societal conflict
small community	minimization of unaccountable power
popular sovereignty	centers
universal adult suffrage	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: Definitions of Authoritarian Regime and a Low Degree of Polyarchy Contrasted**

Authoritarian Regime (Linz 1975, 264):

[Political systems without] free competition between leaders to validate at regular intervals by nonviolent means their claim to rule. . . .<sup>39</sup>

. . . political systems 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.

Polyarchy Scale Score 5 (Coppedge and Reinicke 1990, 53-54):

[There are] no meaningful elections: elections without choice of candidates or parties, or no elections at all.

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.

**Figure 1: Intension and Extension of Linz's Definitions of Regime Types**

ideology of leaders	participation	institutionalized?	full pluralism		limited pluralism		monistic control	
			elections	not	elections	not	elections	not
indeterminate 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						

### Notes

1. 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.
2. Linz, "Totalitarian and Authoritarian Regimes," p. 264, quoting the definition from his own "An Authoritarian Regime: The Case of Spain" from 1964.
3. Michael Coppedge and Wolfgang Reinicke, "A Scale of Polyarchy," *Studies in Comparative International Development* 25:1 (Spring): 51-72.
4. Robert A. Dahl, *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1970), figure 1.2, p. 7.
5. 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).
6. Fernando Henrique Cardoso and Enzo Faletto, *Dependencia y desarrollo en América Latina* (México: Siglo Veintiuno, 1971).
7. 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.
9. Guillermo O'Donnell, "Horizontal Accountability and New Polyarchies," Kellogg Institute Working Paper No. 253 (April 1998).
10. Ruth Berins Collier and David Collier, *Shaping the Political Arena* (Princeton: Princeton



University Press, 1991).

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

12. 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.

13. Linz wrote hundreds of pages describing political systems that differed from these basic three. Some, such as "authoritarian situations," could fit in Figure 1 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 1 and therefore suggest that Linz's underlying classificatory scheme was still more complex.

14. Technically, the Polyarchy Scale is not continuous, but a set of 11 ordered categories.

However, the principle would be the same for truly continuous indicators even though it would be harder to identify the threshold that is the closest equivalent to the categorical definition.

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

16. 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.

17. 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.
18. Arend Lijphart, *Patterns of Democracy* (New Haven, Conn.: Yale University Press, 1999), pp. 243-50.
19. 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).
20. 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.
21. Robert A. Dahl, *Democracy and Its Critics* (New Haven: Yale University Press, 1989), pp. 24-33.

22. Joseph A. Schumpeter, *Capitalism, Socialism, and Democracy* (New York and London: Harper and Brothers, 1942); Robert A. Dahl and Charles E. Lindblom, *Politics, Economics, and Welfare: Planning and Politico-Economic Systems Resolved into Basic Social Processes* (New York: Harper, 1953); Dahl, *Polyarchy*; Giovanni Sartori, *Democratic Theory* (Westport, Conn.: Greenwood Press, 1973).
23. Seymour Martin Lipset, "Some Social Requisites of Democracy: Economic Development and Political Legitimacy," *American Political Science Review* 53 (March 1959): 69-105.
24. Phillips Cutright, "National Political Development: Measurement and Analysis," *American Sociological Review* 28 (1963): 253-64.
25. Kenneth Bollen, "Political Democracy: Conceptual and Measurement Traps," in Alex Inkeles, ed., *On Measuring Democracy: Its Consequences and Concomitants*, New Brunswick: Transaction, 1991), pp. 10-12.
26. Tatu Vanhanen, *The Process of Democratization: A Comparative Study of 147 States, 1980-88* (New York: Crane Russak., 1990).
27. Freedom House "Survey Methodology," in *Freedom in the World 1990-91* (New York: Freedom House, 1991) pp. 49-51.
28. Dahl, *Polyarchy*.
29. Coppedge and Reinicke, "A Scale of Polyarchy."
30. Kenneth A. Bollen, "Issues in the Comparative Measurement of Political Democracy," *American Sociological Review* 45:3 (June 1980): 370-390.
31. Keith Jagers and Ted Robert Gurr, "Tracking Democracy's Third Wave with the Polity III Data," *Journal of Peace Research* 32 (November 1995): 469-82.
32. Axel Hadenius, *Democracy and Development* (Cambridge: Cambridge University Press,

1992).

33. Bollen, "Liberal Democracy," p. 1218. 1207-30.

34. Alex Inkeles, "Introduction: On Measuring Democracy." *Studies in Comparative International Development* 25:1 (Spring 1990): 5-6.

35. Bollen, "Liberal Democracy," p. 1218.

36. Inkeles, "On Measuring Democracy," 5; Gerald W. Hoppie and Jo L. Husbands, eds., "Assessing Progress Toward Democracy: Summary Report of a Workshop," Panel on Issues in Democratization, Commission on Behavioral and Social Sciences and Education, National Research Council (Washington, DC: National Academy Press, 1991), pp. 11-12.

37. A partial exception is Axel Hadenius, who made a start by combining indicators of contestation and participation in an innovative and promising fashion (Hadenius 1992).

38. S.S. Stevens, "On the Theory of Scales of Measurement," *Science* 103 (1946): 677-680.

39. 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).