4

Thickening Thin Concepts: Issues in Large-N Data Generation

Michael Coppedge

This chapter introduces the distinction between 'thick' and 'thin' approaches to data generation and spells out the trade-offs that are associated with these two conventional approaches in comparative politics. I discuss these two approaches as they relate to matters of conceptualization and measurement and identify trade-offs between generality and specificity, quantity and quality, and absolutes and matters of degree. I illustrate these principles with references to research on democracy.

The trade-offs I address recur frequently in the literature and thus scholars are by now familiar with them. Yet there is no reason inherent in either approach for such trade-offs to exist. The trade-offs have been imposed on political scientists by practical limitations, especially the scarcity of appropriate data. To be sure, it will take an enormous effort to collect more data, different data, and better data than we currently have. But the chasm between these approaches can and should be bridged.

4.1. 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. In contrast, thin concepts tend to be simple, unidimensional, and more theoretically adaptable. These differences have important implications. Indeed, some highly consequential trade-offs are associated with these differences.

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.

Table 4.1. Elements of Held's model of democracy

Regular elections	Universal adult suffrage
Elections for many offices	Proportional representation
Secret ballot	Independent, professional bureaucracy
Public debates	Unbiased state
Jury service	State with interests of its own
Strong executive	Individualism
Strong leadership	Representation of corporate interests
Party politics	Private property
One person, one vote	Market economy
Multiple or different voting rights	Patriarchal family or society
Representation	Large nation-state
Constitutional limits to state power	Global state
Separation of powers/checks and balances	International competition
Participation in local government	Professional bureaucracy
Guarantees of civil liberties	Experiments with collective property
Guarantees of political rights	Industrial society
Competition for power	Nonindustrial society
Interest-group pluralism	Poorly informed or emotional voters
Autonomous civil society	Culture of toleration
Rule of law	Economic inequality
Free-market society	Procedural consensus
Workplace democracy	Consensus on legitimate scope of politics
Internal party democracy	Priority of economic interests
Openness to institutional reform	Moderate level of participation
Transparency	Exclusion of some from effective
Representation of the powerful	participation by economic inequalities
Mixed government	Pluralist, free-market international order
Direct participation	Unequal international order
in decision-making	Liberal leadership
Some appointments by lot	Limited bureaucracy
No distinction between citizens	Restriction of some interest groups
and officials	Redistribution of resources
Strict term-limits	Right to childcare
Payment for participation	Maintenance of religious worship
Public campaign finance	Intense societal conflict
Innovative feedback mechanisms	Minimization of unaccountable
Small community	power centers
Popular sovereignty	Demilitarization

Source: Author's compilation of elements discussed in Held (1996).

A thick version can refer to dozens. For example, David Held's *Models of Democracy* (1996) defines twelve different models of democracy, all of which, he argues, possess some claim to the democratic label. Between them, these twelve models refer to seventy-two different characteristics, which are listed in Table 4.1. 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 (1975) of an authoritarian regime:

> [Political systems without] free competition between leaders to validate at regular intervals by nonviolent means their claim to rule...[1] 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.[2]

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 4.2.) 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 4.2.) 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 (Coppedge and Reinicke 1990). 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 to be the case 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

---

[1] This element is implied by Linz's explicit statements that authoritarian regimes are by definition nondemocratic. The language comes from Linz's own definition (1975: 182–3) of a democratic political system.

[2] Linz (1975: 264), quoting the definition from his own 'An Authoritarian Regime: The Case of Spain' (Linz 1964).

Thickening Thin Concepts

Table 4.2. Definitions of authoritarian regime and a low degree of polyarchy contrasted

Authoritarian regime (Linz 1975: 264):	Polyarchy Scale score 5 (Coppedge and Reinicke 1990: 53–4)
[Political systems without] free competition between leaders to validate at regular intervals by nonviolent means their claim to rule.*	[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.	

* This element is implied by Linz's explicit statements that authoritarian regimes are by definition non-democratic. The language comes from his own definition of a democratic political system (Linz 1975: 182–3).

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 × 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. Robert 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 (Dahl 1971: 7).

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 (Lakatos 1978). 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 (Cardoso and Faletto 1979). Guillermo O'Donnell (1973, 1999*a*: ch. 8, 1999*b*) has formulated a series of others—bureaucratic-authoritarianism, delegative democracy, and horizontal accountability. The Colliers' 'mode of incorporation' is yet another (Collier and Collier 1991). Some thick concepts would qualify as 'conflicting imperatives', Andrew Gould's term (1999) for complex concepts possessing a tension that can be used to generate hypotheses. All of these could be considered either very thick concepts or shorthand for theories.

One of the trade-offs between thick and thin concepts was spelled out long ago by Giovanni Sartori (1970).[3] 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 (2001) offers a fine example in his discussion of the concept of '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 4.1 illustrates this trade-off using Linz's definitions (1975) 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 mind-set of the leaders, and the degree to which the political system is 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 3 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.[4] 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

---

[3] See Collier and Levitsky (1997) for variations on this theme.

[4] Linz wrote hundreds of pages describing political systems that differed from these basic three. Some, such as 'authoritarian situations' (Linz 1973) could fit in Figure 4.1 without any revisions to the characteristics around which it is structured. Others, such as sultanistic regimes and post-totalitarian regimes (Linz and Stepan 1996: ch. 3; Chehabi and Linz 1998*b*), had defining characteristics that were not part of Figure 4.1 and therefore suggest that Linz's underlying classificatory scheme is still more complex.

Ideology of leaders	Participation	Institutionali-zation	Pluralism					
			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						

Figure 4.1. Intension and extension of Linz's definitions of regime types

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 mind-set.

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.[5] 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.

4.2. THICKENING THE CONCEPT OF DEMOCRACY

Attempts to redefine 'democracy' illustrate these trade-offs. 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

---

[5] Technically, the Polyarchy Scale is not continuous, but a set of eleven ordered categories. However, the principle would be the same for truly continuous indicators even though it would be harder to identify the threshold; i.e. the closest equivalent to the categorical definition.

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 fifty years and quite understandably call for them to be restored or made explicit. Thus Philippe Schmitter and Terry Karl (1991) include institutionalization and a viable civil society ('cooperation and deliberation via autonomous group activity') among their criteria for 'what democracy is'. Similarly, others stress the centrality of the rule of law and an independent judiciary (Hartlyn and Valenzuela 1994; Diamond 1999: 111–12; O'Donnell 1999*a*: ch. 8). Samuel 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 (Valenzuela 1992: 62–8; O'Donnell 1999*a*: ch. 7).

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 4.1: 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 (1971) 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 decision-making—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 decision-making: 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. Some 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 Arend Lijphart's concept (1999: 243–50) 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. 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 decision-making 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.[6] 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.

4.3. 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 was '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.

---

[6] 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 Proyecto Estado de la Nación (2001) and O'Donnell (2004).

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 threefold 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. Moreover, 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 Beethoven, for example, 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 0 or 1. 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 David Collier and Robert Adcock (1999), 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, that is, 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.[7]

4.4. MEASURING DEMOCRACY

Existing indicators of democracy are just beginning to satisfy rigorous standards for measurement. Democratic theorists before 1776 first simplified the

---

[7] 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, e.g. '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.

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 (Dahl 1989: 24–33). 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 World War II 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 (Schumpeter 1942; Dahl and Lindblom 1953; Sartori 1962; Dahl 1971). 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, Seymour Lipset (1959) 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). Phillip Cutright's index (1963) of 'national political development' was the sum of a country's democracy scores over a twenty-one-year period, which made the number of years of democracy matter as much as the degree of democracy in each year. As Kenneth Bollen (1991: 10–12) has observed, this mistake has been repeated several times, even as late as 1988. 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, Tatu Vanhanen (1990) included the percentage of the vote won by 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. 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 (Freedom House 1991). 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 (Dahl 1971). 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. Michael Coppedge and Wolfgang Reinicke (1988) 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. Many of the existing quantitative indicators of 'democracy' are actually indicators of contestation. They are the Bollen (1980) index of Political Democracy, the Polity data on democracy and autocracy (Jaggers and Gurr 1995), the Freedom House ratings of Political Rights and Civil Liberties (Piano and Puddington 2005), the Polyarchy Scale (Coppedge and Reinicke 1990), Axel Hadenius' (1992) Index of Democracy, and Bollen's Index of Liberal Democracy (1993). It has been demonstrated repeatedly that these indicators measure the same underlying dimension. Their intercorrelations, for example, usually exceed 0.83 (Inkeles 1990).[8]

The indicators we have are by no means perfect: Bollen (1993: 1218) has demonstrated, for example, that Freedom House ratings, at least for 1979–81, tended to underrate Eastern European countries and overrate Latin American countries by a small but statistically significant amount. 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

---

[8] For an excellent survey and evaluation of the existing indicators, see Munck and Verkuilen (2002).

a point measure; only a few are time series, and unfortunately, as Gerardo Munck and Jay Verkuilen (2002) 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) data-set (Przeworski et al. 2000), the Freedom House indices or the Polity index. 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 (1993), the Freedom House Political Rights ratings for 1979–81 were 93 percent valid despite the regional bias. It also correlates at 0.938 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 (Inkeles 1990: 5). Of course, as Munck and Verkuilen (2002) 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 three-point scale rather than a ten-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 three-point scale of democracies, semidemocracies, and nondemocracies—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 'nondemocracy'. 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 (Collier and Adcock 1999; Elkins 2000).

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-à-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.[9] Doing so requires a stronger theory about how dimensions of democracy combine, from which one might derive a mathematical formula. Munck and Verkuilen (2002) 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 (Stevens 1946). Most measurement of democracy now is ordinal, so if we wish to develop a single indicator of democracy in several dimensions, we

---

[9] Two partial exceptions, which make a start by combining indicators of contestation and participation in an innovative and promising fashion, are Axel Hadenius (1992) and the Electoral Democracy Index that Gerardo Munck constructed for the UNDP report *Democracy in Latin America* (UNDP 2005: 21–33 of statistical compendium).

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.

4.5. CONCLUSION

We are far from creating all the rich data that would be needed to measure any thick concept of democracy in a large sample. Indeed, 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.