Much of our knowledge about politics is factual or descriptive. It requires useful concepts and valid, reliable measurement, as discussed in chapter 2. Often, it goes beyond describing static, unchanging situations and constructs narrative accounts of specific events. These accounts can be richly detailed, amounting to histories of events and processes that enable us to feel that we understand what happened and why. Much of our understanding of the birth and death of democracies consists of this kind of knowledge. If one wants to understand how democracy evolved in Great Britain, there are hundreds of books that recount the process in exhaustive detail. Most other countries are not quite as well studied, but there are still dozens of books on democratization in Brazil and Japan and even a healthy number of books or articles on relatively neglected cases such as Paraguay, Botswana, Sri Lanka, and Mongolia. As I argue below, the knowledge contained in these works is essential and extremely valuable; it is a great scholarly achievement that has been tragically undervalued.

This literature is so massive that it would be foolish to attempt to summarize it in this book. Instead, I focus here on attempts to develop a theoretical understanding of democratization, which is a quite different kind of knowledge. (Readers can judge for themselves whether it is a superior or inferior kind of knowledge.) Theoretical understanding is knowledge of the general laws that give rise to the specific events we observe. If we had perfect theoretical understanding, we could claim that the causes singled out in our explanations will have the same effects in cases that we have not yet observed, and would have the same effects even in hypothetical situations that we can never observe. We do not have this quality of understanding yet, and we may never have it. But the goal of political science, just like physics or biology or meteorology, is constantly to improve our theoretical
understanding and to learn useful things along the way.

Practitioners of different approaches to comparative politics have jealously disputed one another’s claims to theoretical understanding. Area studies specialists have accused quantitative comparativists of either comparing the incomparable or quantifying the obvious. Statistical analysts have condescendingly thanked country experts for digging up the anecdotal evidence that only multi-country comparisons can transmute into theory. Both complain about the lack of realism in rational-choice theory, yet formal theorists have tried to brand propositions in both large-sample and case-study approaches "atheoretical" because they are not integrated into a larger, systematic body of propositions. All these charges reflect a narrow view of theory. In reality, all three approaches make indispensable contributions to good theorizing.

In this chapter I define three fundamental criteria for good theories, and I use these criteria to evaluate three approaches in comparative politics—formal theory, case studies and small-sample comparisons, and large-sample statistical analysis. My purpose in doing so is to advocate a broader view of theory, in which each approach has one unique strength and two weaknesses. From this broad perspective, our three main approaches can be seen as complementary. I illustrate the tradeoffs with examples from research on democratization, which has been studied so long and so variously that it affords examples of the strengths and weaknesses of every method.

**Three Central Criteria for Good Theory**

An overview of criteria for good theorizing provides a good foundation for a comparison of the advantages and disadvantages of different approaches. In a literature too vast to summarize here, scholars have defined more than a dozen criteria for good theory. However, I contend that three criteria are especially central: generality, integration, and thickness. They are central in the sense that each of the three major approaches in comparative politics achieves one at the expense of the
others. Case studies and small-sample comparisons produce thick propositions, large-sample statistical research produces general propositions, and formal theory integrates propositions. This division of labor has developed because these three criteria are locked in a three-way tradeoff: it is very difficult, in practice, to do a good job of satisfying more than one criterion at a time.

**Generality**

A general theory is one that is intended to apply to all relevant cases, both all those that have been observed and all that could be observed.³ (A general theory must also be correct for all cases, but I leave this discussion to the section on testing, below.) Some scholars claim not to seek general knowledge and consider the goal of generalization a matter of taste. Sir Isaiah Berlin once suggested that people are either foxes, who know many small things, or hedgehogs, who know one big thing.⁴ I think a better analogy for my purposes would contrast whales and octopuses. Both are renowned for their intelligence, but they use their intelligence in different ways. Whales come to know great swaths of the earth in their tours of the globe; they lack limbs that would allow them to experience objects first-hand; and their eyesight is too poor to perceive fine detail. They acquire a surface knowledge of general things. Octopuses, in contrast, dwell in one place and use their fine eyesight and eight infinitely flexible arms to gain an intimate knowledge of local, specific things. (To buttress the analogy, there is the additional, although not apropos, parallel that octopuses are well equipped to blend into their surroundings, while whales are as conspicuous as creatures can be. However, I ask readers not to overinterpret the octopus' tendency to spread clouds of ink when threatened.) I do not wish to suggest that scholars who emulate the octopus should emulate the whale instead, or vice versa. Rather, my point is that each kind of knowledge is limited in its own way and that the most complete kind of knowledge would result from pooling both kinds.

For decades, the proponents of "middle-range theory" have fostered the view that
generalization is, at best, optional, and at worst, impossible and pernicious. This is false. Generality is an indispensable characteristic of theory. In the standard (nomological) view of what theories are, an explanation interprets an event or a tendency as a specific instance of universal laws. If the laws are not universal, then there is no solid foundation for the explanation; the explanation itself then requires explanation, and that explanation requires explanation, and so on. The phenomenon is not explained until it is understood as a necessary consequence of laws recognized as universally true.

Because true universals are unattainable in social science today, practicality forces us to confine our theories to bounded times and places. We must take care not to bound our theories arbitrarily, that is, for no good theoretical reason. But as long as we can presume that there is potentially a good theoretical reason for limiting a theory to, say, postwar Europe, even if the reason is vague or implicit, then we can treat the theory as provisionally valid, pending completion (and empirical confirmation). All actual theories in comparative politics are therefore incomplete and provisional. The admission that they are works in progress is not damning, because this is all that one can claim about any scientific theory. Even physicists are still searching for a Theory of Everything, which could overturn the insights of Einstein and quantum mechanics. It is probably much harder to generalize in comparative politics than it is in physics, but we still have an obligation to generalize, for generalization is a defining characteristic of theory.5

The notion of a truly general theory of democratization in units as diverse as countries is difficult to imagine. Even those who dare to generalize about a region typically take care to disclaim any application beyond that region. After all, it is very likely that the nature and causes of democratization were fundamentally different in 19th-Century Western Europe than they were in Latin America in the 1980s. In Europe, democratization required extending the suffrage to all adults
in elections that were already competitive; in Latin America, it required restoring electoral
competition to countries that had usually already practiced universal adult suffrage. These were
fundamentally different processes involving different sets of actors, interests, and goals. How could
one general theory fit both processes? The answer is that it is fine to explain these different
processes in different ways as long as our general theory includes a reason for knowing where to
apply which part of the theory. A general theory might state, for example, that the European variant
of the theory applies only to the first countries to develop mass liberal democracy, while a different
version of the theory applies to the latecomers, because pro-democracy elites in the latecomers
benefited from the example of the first democracies. Without such a reason, a theory would be
arbitrarily bounded. It would state, in effect, that the theory applies everywhere except where it
doesn’t, or that it applies only in Western Europe because we say so. The justification for bounding
a theory does not have to be fully developed, because general theory is a work in progress; but
scholars have an obligation at least to recognize this issue and to suggest possible reasons for
circumscribing the relevance of their theories to certain times and places.

Integration

Generalization to the entirety of observed reality is not enough. In order to explain, we must
also generalize to what is unobserved and hypothetical. As Donald Moon wrote,

The nomological pattern of explanation, as its name implies, requires the presence of general
laws in any explanatory account. . . . But not just any kind of general statement can perform
this explanatory function. Not only must laws be unrestricted universals (i.e., they must be in
universal form and must apply to an unrestricted class of objects), but they must also
support "counter-to-fact" and subjunctive conditional statements. . . . But to make such an
assertion requires a great deal more information than that conveyed by a particular law, and
so in order to understand the explanatory force of laws, we [must] examine them in relation
to scientific theories.  

In other words, instead of saying that a law is generally true because we have observed it to be
generally true, theory says that a law is generally true because it is necessarily entailed by other laws
that are generally true. (Those other laws are in turn subject to the same standards of confirmation.)
Thus a generalization must be integrated into a systematic set of other laws—a theory—in order to be
truly general.

Two qualities and two consequences of this integration deserve to be highlighted. The two
qualities are logical consistency and internal completeness. The propositions of a theory are logically
consistent when they do not contradict one another. Consistency can refer either to the internal
consistency of the propositions in just the theory at hand, or to external consistency with
propositions in other theories. A theory is internally complete when it defines a prediction for every
logical-entailed premise: a “then” for every “if.” Theories that raise a lot of “ifs” but predict few
“thens” are incomplete. For examples, regime typologies could be considered incomplete theories.
They highlight selective characteristics of regimes, such as elections, competition, and freedom, but
they do not specify what difference it makes whether a regime possesses these characteristics or not.
A more complete theory would tell us that if a regime is democratic, then something specific will
result, such as extensive social security coverage or peaceful relations with other democracies.

If a theory is internally complete and internally and externally consistent, then two highly
desirable consequences follow. First, consistency makes it possible for a theoretical understanding to
cumulate. When the propositions of one theory are logically consistent with one another and with
the propositions of another theory, the two theories are integrated with each other, merging into a
larger, more complex and comprehensive theory that attempts to explain more of the world. As
John Gerring notes, "The proposition that sits by itself in a corner is likely to be dismissed as 'ad hoc,' or 'idiosyncratic.' It does not fit with our present understanding of the world. It refuses to cumulate." The more ways in which a proposition meshes with other propositions, the richer our understanding becomes.

Second, consistent and complete theories are more “fertile”: they generate a larger number of observable implications. This results in part from the completeness of the theory, but it is also a by-product of the number of propositions that are integrated together. The greater the number of propositions that are linked together, the more hypotheses they can generate. One can derive more theorems from fifty axioms than from three. As a result, complete, well-integrated theories have many observable implications and are therefore potentially more testable.

The basis for the systematic structure of a theory is often logic, but it can be other branches of mathematics as well, such as calculus, game theory, or probability theory. I believe that it can also be, and typically and unavoidably is, common sense: our own informal understandings of how the world works. No elaborate proof is needed to show that money talks, that united organizations are stronger than divided ones, or that you can fool some of the people some of the time. These understandings of how the world works are less foolproof than mathematical or logical tools, but consciously or not, we rely on them all the time. For example, if a general calls for the overthrow of an elected president, we may not know exactly what will happen, but the range of possible consequences is actually quite small. The general may be forced to retire or sent overseas, other officers may rally around him, the U.S. ambassador will weigh in or one side or another, and so on; but we know the consequences will not include the discovery of a platinum mine, a major earthquake, or the outbreak of world peace and harmony. Our common sense guides the translation of theoretical symbols into meaningful referents (interpretive theory) and informs and constrains the
range of possible causal connections (causal theory). In fact, few hypotheses in comparative politics have been derived purely from the formal assumptions of a theory. In almost all cases, at some point, researchers have had to draw on their common-sense knowledge of the political world to translate the logical implications of a theory into observable implications.

**Thickness**

Finally, theory should be thick. A “thick” theory is a thorough one, a theory that provides a complete explanation for the phenomenon in question. It is useful to think of theoretical thickness as having two dimensions: depth and breadth. A theory is “deep” if it traces the chain of causation far back from the eventual effect. Depth is desirable to avoid overly proximate “explanations,” which tend to be superficial or trivial. For example, Higley and Burton argued that “a disunified national elite... produces a series of unstable regimes that tend to oscillate between authoritarian and democratic forms...” while “a consensually unified national elite... produces a stable regime that may evolve into a modern democracy...” Although their argument fit their cases well, the authors never explained why a country’s elite is divided or consensually unified. The cause is suspiciously close to the effect, so the explanation is unsatisfying. It avoids the more interesting, and more difficult, question of what causes elite unity or disunity. A deeper explanation that took us farther back along the causal chain would be more useful and satisfying.

The breadth or complexity of a theory concerns the number of parameters it includes and the degree of interconnection among them. Every theoretical model in social science has five parameters. First, every model pertains to a certain level of analysis--individual, group, national, world-systemic, or some intermediate gradation. Second, it has one or more dependent variables. Third, it has one or more explanatory variables. Fourth, it applies to a certain relevant universe of cases. And fifth, it applies to events or processes that take place during a certain period of time. We
can refer to the definitions of each of these five parameters as possessing "zero-order complexity" because no relationships among parameters are involved. In the study of democratization, however, even at the zero order there is great leeway for defining what democracy is, how to measure it and any explanatory factors, which sample of countries is relevant for testing any given set of explanations, and the period of time to which such explanations apply. And this is just at the national level of analysis; with smaller or larger units of analysis, one would use completely different variables, cases, and time frames.

"First-order complexity" involves any causal relationship between any of these parameters and itself. These relationships include:

1. causation bridging levels of analysis, or aggregation and disaggregation;
2. causal relationships among dependent variables, or endogeneity;
3. interactions among independent variables;
4. impacts of one time period on another, called lagged effects or temporal autocorrelation; and
5. the impact of one case on another, called diffusion or spatial autocorrelation.

Such relationships may sound overly technical and irrelevant, but in reality, examples of all of them can be found in the democratization literature. O'Donnell and Schmitter proposed aggregation in the theory of democratization at the national level as the outcome of strategic maneuvering among elites at the group or individual level. Also, their argument that political liberalization is a prerequisite for transition asserted an endogenous relationship between these two outcomes. Anyone who has studied modernization as a cause of democracy assumes that the components of modernization, such as education, wealth, urbanization, and secularization, interact to produce democracy. Many quantitative studies assume that democratization is a process of incremental change from a country's previous level of freedom. These are lagged effects. Finally, a
number of recent studies have examined the democratic diffusion hypothesis that conditions in other countries influence democratization.\textsuperscript{13} First-order complexity is common.

Second-order complexity involves causal relationships between two different parameters. All hypotheses about an independent variable causing democracy (or democracy causing something else) are of this order; but so are various complications that could be introduced into a model. If the meaning of democracy varies over time or the best way to operationalize an independent variable depends on the world region, then one is dealing with this degree of complexity. Third-order complexity comes into play when there are plausible hypotheses relating three parameters. Most common among these are hypotheses that the relationship between the dependent variables and an independent variable is partly a function of time or place. A good example is the hypothesis that the impact of economic development on democratization depends on a country's world-system position.\textsuperscript{14} With fourth-order complexity, a causal relationship could be a function of both time and place (or level of analysis). This may sound far-fetched, but in small-sample comparisons such relationships are fairly commonly asserted--for example, the notion that increasing wealth has not favored democracy in the Arab oil-producing states since the Second World War;\textsuperscript{15} or the claim that the U.S. has become more sincerely interested in promoting democracy in the Caribbean Basin since the end of the Cold War.\textsuperscript{16} Increasing complexity does not render a theory more esoteric; on the contrary, it is only by increasing complexity that a theory begins to approximate common sense.

Orders of complexity can increase only so far. Eventually, one arrives at the extremely inelegant "saturated" model that explains each outcome perfectly by providing different and unique explanations for each case. Laypersons who have not been socialized into social science know that the saturated model is the truth: every country is unique, history never repeats itself exactly, and every event is the product of a long and densely tangled chain of causation stretching back to the
beginning of time. We political scientists know on some level that a true and complete explanation for the things that fascinate us would be impossibly complex. But we willfully ignore this disturbing fact and persist in our research. We are a community of eccentrics who share the delusion that politics is simpler than it appears. Although I would be as delighted as any other political scientist to discover simple, elegant, and powerful explanations, I think the common sense of the layperson is correct: we must presume that politics is extremely complex, and the burden of proof rests on those who claim that it is not.

Guided by our own experience in the world, we should presume that most of these complex possibilities could be true and that only a complex theory can explain politics well. Unfortunately, this is a controversial position. Most influential works on the methodology of comparative politics emphasize the fact that all models necessarily simplify reality, and these texts usually exalt parsimony as a methodological virtue. However, parsimony is frequently misunderstood. It is not a rule that we should always prefer the simpler of two theories; properly understood, it is a rule that if two theories explain a phenomenon equally well, we should prefer the simpler one. In my experience in comparative politics, simplifications almost always sacrifice some accuracy. I see no reason to prefer a simple but less accurate theory over a complex but more accurate one. Of course, theoretical thickness does not guarantee a grasp of the truth; any creative person could dream up ten complex theories that are wrong for every one that is right. But very few of the right theories are likely to be simple. We should not let a misguided preference for parsimony blind us to the truth. We have to consider complex theories; the trick is to find the right ones. This is the role of testing, which will be discussed in chapter 7.

Multiple Paths to Theory

All approaches in comparative politics are deficient in satisfying some requirements for
theory. In fact, each of the three major approaches excels in a different respect and is deficient on the other two. This is why they are competing approaches. Formal theory integrates propositions into a larger web of theory but neglects generalization and is thin; large-sample statistical analysis establishes general empirical fit, but in thin and often rather ad hoc ways; and case studies and small-sample comparisons generate thick knowledge that may not be generally true and tends to be only loosely integrated into a larger theory. The following sections elaborate on this evaluation of each approach according to these three criteria, with illustrations from research on democratization.

Theory in Small-Sample Studies

Case studies and small-sample comparisons sometimes have been dismissed as "merely" descriptive, anecdotal, historical, or journalistic, and therefore atheoretical. But the harshest critic of such studies would have to concede that they at least generate "facts." Facts may seem atheoretical, but they are not. In reality, we cannot even observe facts until we have a conceptual framework to help us make sense of what we perceive. Such a framework is an interpretive theory that helps us identify what to observe, defines the relevant and meaningful characteristics of actors and institutions, and fills in the connections between action and reaction so that we can plausibly reconstruct events and processes. We are unconscious of much of this theory; we develop it and refine it by trial and error from birth onwards. If the test of a good theory is that it enables us to predict consequences, common sense is a magnificent theory. With it, we successfully walk, talk, drive, work, parent, and invest, negotiating our way around a thousand daily challenges throughout whole lifetimes. With the benefit of this common-sense understanding of how the world works, we feel that we can understand political events that we did not experience in person, if only someone will supply us with the crucial details.

The more descriptive case studies and small-sample comparisons consist of propositions that
are integrated into this intuitive interpretive theory. The bulk of democratization research consists of case studies and small-sample (usually within-region) comparisons. Every transition to democracy in the past two decades has been thoroughly analyzed in several books and numerous articles. Some of the most influential books in the field have been compendia of case studies. Scholars seeking a thorough knowledge of a particular transition, breakdown, or regime survival are practically cursed with a superabundance of information. Often such studies prefer specific concepts to general ones: Duma to parliament, Clinton to president, Chiapas to province; but such precision reflects not a simple interpretive theory, but a more elaborate one that captures some of the idiosyncracies of each case.

What is striking at this level is that we collectively know so much and disagree so little. Research of this type has created what is probably the most thorough understanding of specific democratic transitions, breakdown, and survival, and has done so for practically every country one could mention. These works, whether they are academic research in a British or anthropological tradition, current history, or journalistic analyses, do an excellent, often unsurpassed, job of recounting events, identifying key actors and their motives, assessing the strength of organizations, and tracing connections among causal forces. The authority of this work is such that we rarely argue about who the key players were, what the basic chronology was, or who won and who lost. Ironically, the lack of controversy about these inferences diminishes the prestige of the scholars who make them. But the high degree of consensus around their work makes their accomplishment more impressive, not less so. All theories should be as convincing as these.

But these studies are just one pole of a continuum in small-sample research. At the opposite pole, some small-sample studies interpret every specific actor, institution, trend, and situation as a specific instance of a general type. They take literally Przeworski and Teune's call to "replace proper
names of social systems” with “the relevant variables.” The kind of theory generated by this type of research tends to have two characteristics. First, most of it is qualitative and categorical. The causal relationships it identifies link types to types and kinds to kinds rather than matching quantities or degrees. Relationships are hypothesized to be true or false, necessary or sufficient, rather than partially true, stronger or weaker, likely or iffy. This qualitative bent does not make this style of theorizing inferior; rather, it is merely different from a mathematical style.

Second, the theoretical propositions that emerge from these studies, if examined with care, turn out to possess a high order of complexity. The more faithfully a theory represents our complex world, the more complex it must be. (How faithfully, of course, is a question to be resolved by testing.) In the Latin American democratization literature, the conventional wisdom presumes that each wave of democratization is different, that each country has derived different lessons from its distinct political and economic history; that corporate actors vary greatly in power and tactics from country to country, and that both individual politicians and international actors can have a decisive impact on the outcome. This is the stuff of thick theory, and comparative politics as a whole benefits when a regional specialization generates such rich possibilities.

For these two reasons, case and area studies have made many of the best-known and most original contributions to comparative political theory. Dependency theory germinated in a study of Argentina's terms of trade. Consociational democracy was inspired by Lijphart’s Dutch origins. The debate about the impact of parliamentary and presidential constitutions began as an effort to understand the fall of the Weimar Republic and its renewal was inspired by the success of the Spanish transition.

The hypotheses generated by this literature have reflected high-order, complex theorizing. Daniel Lerner’s seminal work on modernization was a case study of Turkey that identified parallel
trends in the complex processes of urbanization, secularization, and education.\textsuperscript{25} Juan Linz’s theorizing about the breakdown of democratic regimes described a detailed sequence of events–crisis, growing belief in the ineffectiveness of the democratic regime, overpromising by semiloyal leaders, polarization of public opinion, irresponsible behavior by democratic elites, culminating in either breakdown or reequilibration. He saw each step as necessary but not sufficient for the next, and described various options available to elites at each stage, as well as structural and historical conditions that made certain options more or less likely. This was a theory that assumed endogeneity, aggregation across levels of analysis, and conditional interactions among causal factors.\textsuperscript{26} O’Donnell and Schmitter bridged levels of analysis when they theorized about democratization at the national level as the outcome of strategic maneuvering among elites at the group or individual level; they contemplated endogeneity or path dependence when they asserted that political liberalization was a prerequisite for regime transition.\textsuperscript{27} Huntington’s thesis that there are waves of democratization required a transnational causal process in addition to multiple domestic causes.\textsuperscript{28} The Colliers’ \textit{Shaping the Political Arena} identified four similar processes or periods–reform, incorporation, aftermath, and heritage--in eight cases but allowed them to start and end at different times in each country. It was particularly exacting in describing the nature of oligarchic states, organized labor, and political parties and in specifying how they interacted with one another, and with many other aspects of their political contexts in the 20th century, to affect the course of democratization.\textsuperscript{29} Case studies of democratization, such as those collected in the Diamond, Linz, and Lipset projects and dozens of country monographs, weave together social, economic, cultural, institutional, and often transnational causes into coherent, case-specific narratives.\textsuperscript{30} This literature has been the source of most of what we think we understand about democratization.
Nevertheless, the small-sample approach has two weaknesses. First, although its propositions are integrated with theory, they are integrated more loosely. By "loosely," I mean that such propositions are not derived from other propositions according to any strict logic. Rather, they are borrowed from other theories and taken out of their original theoretical context or generated by observation, induction, and intuition. Loose integration has two consequences. One is that the facts can be used to support an embarrassing variety of theories. This happens because the question, "What is this a case of?" has many possible answers. The leap from specific to general can go in many different directions. What, for example, was Venezuela in 1989 a case of? Every theoretical framework suggests a different direction. To a progressive political economist, it was an oil-dependent economy; to an institutionalist, it was a presidential partyarchy; to a liberal political economist, a case of delayed structural adjustment; to a student of labor, it was a corporatist system; to a cultural theorist, a nation with unrealistic trust in a "magical state." In reality, all of these labels may have been accurate. The point is that moving from the specific to the general forces us to describe our cases more selectively, and we make our selections so as to integrate the case into a larger body of theory.

The second consequence of loose theoretical integration is that it is less clear which tests would confirm or disconfirm the theory. Without rigorous logic or mathematical tools to generate hypotheses, there is no straightforward way to derive necessary implications: what must be true if the theory is true. In contrast to formal theory, the theories of small-sample analysis are less clear about their assumptions; they rely more on the tacit assumptions of common sense, which leads to conditional and vaguely probabilistic predictions, which are hard to falsify.

The second weakness of small-sample theories is that they are, by definition, not general. These propositions (when they are explicitly integrated into a theory) merely assert generality;
whether such assertions are empirically valid or not is a matter for large-sample testing to decide. Until the testing takes place, these are only general hypotheses, not generally confirmed theory. Replacing proper names with variables is indeed our goal, but generalizing is far harder than affixing general labels to particulars. It is one thing to call the United States a presidential democracy, but quite another to assert that what one observes in the United States is true of presidential democracies in general. The former is a description of one case; the latter is an inference about a population (all presidential democracies) from one case (the United States), which is not justified.

To summarize, case studies and small-sample comparisons yield a type of theory that is qualitatively thick and empirically well-grounded, and therefore plausible in bounded times and places; but also provisional, pending extension to more general samples; and often ambiguous in its theoretical implications, and therefore difficult to test decisively, especially beyond its original boundaries. It is, to caricature a bit, a soft theory built on a hard foundation.

Brian Downing's *The Military Revolution and Political Change* is a small-sample comparison that illustrates this tradeoff well. As we have come to expect of comparative historical analysis, it is admirably thick and well-grounded. Downing delves deeply into the resurrection of Roman Law in medieval constitutionalism, the rights of towns in the Holy Roman Empire, the terms of feudal levies, the advantages of Swiss pikemen over mounted knights, the powers of the estates of Brandenburg and Pomerania, 17th-Century French tax farming, the consensus voting rule in the Polish Seym, the financing of English wars, and the logistics of Swedish armies. The specificity of his arguments, however, forces him to refrain from generalizing beyond Europe:

To say that European social, political, and economic history is markedly different from that of the rest of the world is to say nothing new. . . . Three principal conditions in medieval Europe provided a predisposition to democracy: a rough balance between crown and
nobility, decentralized military systems, and peasant property rights and reciprocal ties to the landlord. Though one or more of these may have obtained in other parts of the world, the combination of all three, as well as the strength of each, was unique to Western Europe.36

It is also difficult to imagine the testable hypotheses that might spring from his theory. A condensed version of his theory is that 1) late-medieval institutions unique to Europe predisposed the region to eventual democratization, but 2) absolutist rulers destroyed these institutions to keep up with an arms race in the 17th Century, although 3) absolutism was avoided in states that were protected by mountains (Switzerland) or water (England), or that could finance the military build-up through trade (Holland). Testing would be difficult in part because the theory is confined to Europe, where there are few cases left unexamined. Testing would also be difficult because the concepts on which the theory rests are so specific that considerable interpretation would be required to apply them to additional cases. Was the country’s late-medieval constitutionalism of the right variety? Was the country under greater military threat than England? Did it earn enough from trade to expand its standing army without raising taxes? One gets the feeling that any apparent exception could be explained away by going more deeply into some obscure and unique 16th-Century land tenure arrangement. Theories based on case studies and small-sample comparisons invite constant reformulation and ad hoc defenses. Nevertheless, they are theories. In fact, they are superior to other theories with respect to thickness.

Theory in Large-Sample Comparisons

Many scholars tend to view large-sample, statistical research an exercise in testing only, rather than a source of theoretical innovation. But even though the original motivation for applying statistics to comparative politics may have been to test hypotheses generated by other methods, this kind of research actually does contribute to theory in distinct and novel ways. The mathematical
tools used in hypothesis testing encourage, and sometimes require, conversion of theories from a qualitative logic to a quantitative logic. Theories become less about kinds and types and true/false or necessary/sufficient relations and more about the magnitudes of impacts, partial impacts, probabilities, and curvilinear relationships. These relationships are difficult to handle in a qualitative idiom. The reverse is not true, fortunately. Statistical analysis can also handle the kinds of relationships found in qualitative theories, such as conditional relations, necessary or sufficient conditions, and the direction of causation.

Examples of distinctly quantitative theory abound in democratization research. The qualitative hypothesis that wealthy countries tend to be democracies has been converted into a rococo variety of quantitative hypotheses:

1. the wealthier the country is, the more democratic it is;
2. the wealthier the country is, the more democratic it is, but to a logarithmically diminishing degree;
3. the wealthier the country is, the more democratic it is, but with logarithmically diminishing increases and a dip at an intermediate level of wealth (the “N-curve” hypothesis);
4. the wealthier the country is, the more democratic it is, except when economic growth temporarily worsens inequality, which undermines democracy;
5. the wealthier the country is, the more democratic it is, although the impact of wealth is mediated by state size, which has an “inverted U” relationship with democracy;
6. increasing wealth does not make countries become more democratic but improves the life expectancy of any regime;

and so on.

Another line of research has begun to explore the notion of democratic diffusion. Although
Rustow and Huntington wrote about various possible types of transnational influences on democratization, quantitative scholars have found that “democratic diffusion” can refer to a tremendous variety of causal paths. In the course of testing for them, they have had to refine the theory in order to distinguish among neighbor effects, regional effects, and superpower effects; impacts on the probability of change, change vs. stasis, the direction of change, and the magnitude of change; and change influenced by ideas, trade, investment, population movement, military pressure, and national reputations, many of which were not contemplated in smaller-sample or qualitative research.

The principal advantage of the kind of theory that emerges from large-sample work is that it is relatively general, both in its aspirations and in its empirical grounding. The degree to which it is general varies depending on the coverage of the universe by the sample, of course, but it is by definition more general than small-sample research. Formal theory makes universalistic assumptions, which are even more general, but large-sample research has the advantage of making at least some assumptions that are guaranteed to have empirical support. (The assumptions of statistical analysis are rarely fully supported, such as the assumption of normally distributed random errors. I will discuss the consequences of this problem in later chapters on testing.) For example, the most consistent finding in the large-sample statistical literature is that democracy is associated with high levels of economic development. The association is a rough one, not strong enough to predict small differences in democracy or differences between any two cases with a high degree of certainty; but it remains a very general statement.

The two weaknesses of large-sample comparisons are thinness and loose theoretical integration. A "thin" proposition is a simple statement that assumes very little about the world and identifies an association between just two narrowly conceived phenomena, such as democracy and
development. Both could be, and originally were, thick concepts that would require thick theory. But large-sample research typically has reduced the concept of democracy to a few easily-measured institutions--fair and competitive elections, some basic freedoms--that reflect just one dimension of democracy: Dahl's notion of "contestation." Similarly, "economic development" has been reduced in this research to per capita GNP, GDP, or energy consumption. Thin concepts make for thin theory. Although the bivariate relationship between thin development and thin democracy has undergone elaborate permutations in statistical testing, many other hypotheses about the causes of democracy have been neglected. None of the large-sample literature really addresses theories that are cast at a subnational level of analysis, such as the very influential O'Donnell-Schmitter-Whitehead project. Large-sample research concerns the national, and occasionally international, levels of analysis, and it will continue to do so until subnational data are collected systematically--an enterprise that has barely begun. In addition, there are quite a few hypotheses about causes of democratization that have not yet been addressed in large-sample research. Among them are U.S. support for democracy or authoritarian governments, relations between the party in power and elite interests, the mode of incorporation of the working class, interactions with different historical periods, U.S. military training, and elite strategies in response to crisis. In this sense, the large-sample literature lags behind the theories developed in other approaches.

The second weakness is loose integration with a body of systematic theory. Mathematics is an extremely systematic tool but by itself it has no political content, and the political content that has been inserted into the mathematical framework lacks system. Large-sample theory consists of a handful of isolated, disembodied propositions. Each one by itself can generate many necessary implications by simply plugging in different numbers. But there is no theory in the gaps between the propositions that would enable us to combine them to make predictions. For example, we know
that rich countries tend to be democratic, and that countries tend to become more like their neighbors. But there is no overarching theory that would enable us to predict how democratic a country should be if it is poor and surrounded by democracies, or rich and surrounded by authoritarian regimes. Lacking such a theory, quantitative researchers tend simply to accept whatever estimates they obtain as reflections of the true weights of different causal factors; and these estimates are then absorbed into the conventional wisdom about what matters. It is a fundamentally inductive process of discovery. The problem with induction, as Hume observed centuries ago, is that even if a hypothesis is confirmed in 10,000 cases, there is no guarantee that it will be true in the 10,001st case unless one has a very good theory that predicts that it must.

Most of the large-sample research on democratization has concerned two debates: one about whether the relationship between development and democracy is linear, logarithmic, or some sort of N-curve; and one examining interactions between per capita GDP, economic inequality and democracy. The lack of integration in this area is so great that after nearly forty years of repeated confirmation, a group of scholars was able to make a credible case that the association is spurious. Similar criticisms could be leveled against the emerging evidence for democratic diffusion (the spread of democracy from country to country): we are pretty sure it is happening, but we do not know what the causal mechanisms are.

This quality of knowing things but not knowing why is what inspires the accusation that large-sample research leads to “empirical generalizations” but not to theory. Again, I consider such criticism to be based on excessively narrow criteria for theory, one that privileges integration over general empirical confirmation and thickness. The propositions that spring from large-sample analysis may be thin and disjointed, but they are still theories, albeit of a distinct kind, with the advantage of empirically-based generality.
Formal theory

Formal (rational-choice) theories achieve levels of integration that elude small- and large-sample research. Three standards for good theorizing often touted by formal theorists are universal scope; clear, simple, and explicit assumptions; and the potential to generate testable hypotheses derived from theory. Formal theorists aspire to universal scope by refraining from limiting the applicability of their theories to certain times and places: what is true for one committee is assumed to be true for all committees as long as the assumptions of the model are met. Formal theorists also make their assumptions simple and explicit, which makes it easier for other scholars to follow the logic of the theory and derive the consequences of modifying some assumptions. Due to its deductive method, formal theory also lends itself to the generation of lots of hypotheses, especially about eventual, stable patterns of collective behavior. Because a whole community of scholars follows this logic and works within it, new research builds explicitly on what has gone before. Theory cumulates.

However, the theory that cumulates is far from thick. Formal theories of democratic transitions, for example, narrow their focus to the final stage of the process in which a small set of elites is bargaining about whether to found a democratic regime or not. This is an extremely thin view of democratization. A thick theory would assume less and attempt to explain more. A thick theory would not limit the identities of the actors or the options before them to a fixed menu of choices. Game-theoretic explanations do not tell us how to know which actors are the elites and how they got their seats at the bargaining table. They do not explain where the bargainers' preferences came from or how they might change. They do not account for the number of rounds of bargaining to begin with, or why democracy is one of the options on the table. A thicker theory would offer explanations (whether cultural, institutional, international, social, or economic) for at
least some of these elements; formal theories simply assume them. Formal theory, as currently practiced, has difficulty developing thick explanations because it is anchored at the individual level. It aspires to make predictions about larger groups, but only within very restrictive assumptions about the rules of the game and the preferences of the players. It is the mirror image of small-sample theory: a hard theory built on a soft base. It is difficult to extrapolate from these small settings to macro-phenomena like regime change. Indeed, Barbara Geddes has called on scholars to stop trying to theorize about “big structures, large processes, and huge comparisons,” such as democratization, for the time being.\footnote{48}

Formal theories have universalistic aspirations; in this sense, they are even more general than large-sample theories, which are confined to the cases actually observed. However, generality is not, or should not be, merely an assertion or assumption of generality. It should be accompanied by empirical confirmation. In this more useful sense, formal theories encounter two obstacles to true generality. First, if taken literally, the assumptions of these models render them inapplicable to the real world. It would not be fair to test hypotheses using a case where the premises from which they sprang did not hold. Yet the premises are typically oversimplifications of reality: that there is a small set of players who perform complex calculations of self-interest with the benefit of full information and no history of past interactions, etc. Such premises cannot be said to hold in real-world situations, and therefore any test of predictions derived from them could be ruled unfair.\footnote{49}

Second, unlike small- and large-sample methods, formal theory is not a method of testing; it is only a method of generating theory.\footnote{50} In a very strict sense, the predictions of formal theories by definition have no empirical confirmation. In a less strict sense, their predictions can be tested; but if they are, they are tested using some version of small- or large-sample research, so the latter provide the only assurances of empirical support. But even if formal theories are generously credited with
any empirical support their predictions may find, it must be confessed that testing is the weak suit of formal theorists. As Green and Shapiro argue, “a large proportion of the theoretical conjectures of rational choice theorists have not been tested empirically. Those tests that have been undertaken have either failed on their own terms or garnered support for other propositions that, on reflection, can only be characterized as banal: they do little more than state existing knowledge in rational choice terminology.” One could object that in practice, formal theorists constantly adjust their assumptions and predictions to what they observe in the world to make it as realistic and relevant as possible. But this back-door induction is so unsystematic that it is prone to all the errors and biases found in the least rigorous testing methods. Nevertheless, the systematic, interlocking, cumulative nature of formal theory is an essential quality of theory, just as thickness and generality are.

Toward Theoretical Perfection

Each approach has one strength and two weaknesses, which complement the strengths and weaknesses of the other two approaches. Formal theory is integrated, but thin and lacking in general empirical support; large-sample analysis is relatively general, but thin and theoretically disjointed; and case studies and small-sample research are admirably thick but drastically bounded and loosely integrated with a systematic set of theoretical propositions. Practical problems have kept these approaches separate in the past, but there is no necessary reason for them to remain separate. Although we can never achieve a perfect theory, we can move in that direction, toward a theory that would be thick, integrated, and general.

Figure 3.1 depicts where we are now and what we would have to do to improve our theories. The figure shows a cube whose three dimensions are the three fundamental criteria for good theory—integration, thickness, and generality. Each dimension ranges from 0 (little achievement of the criterion) to 1 (full achievement of the criterion). I have placed the three major approaches in
comparative politics at an approximate position on these three dimensions. Rational-choice theory occupies a space of high integration, little generality, and thinness. Area studies are associated with thickness but weak integration and little generality. Finally, large-sample statistical analysis occupies a region of high generality but little integration and considerable thinness. At the opposite corner rests our goal: theory that is very thick, fully integrated, and general.

The figure traces two paths from each approach to the unattainable ideal. Formal theory could move toward empirical generalization through testing, which is increasingly common (path 1). Alternatively, formal theorists could thicken their concepts and make their models more complex so that they apply more naturally to real-world situations (path 2). This is one of the aims of the “analytic narratives” approach, which is discussed in chapter 6. Scholars who do case studies and small-sample comparisons could meet them halfway by formalizing the complex models they generate inductively, integrating them with a larger and more systematic body of theory (path 3). But other area-studies scholars will probably prefer to move toward generalization by testing their complex hypotheses in increasingly large samples (path 4). Quantitative comparativists, who have been testing hypotheses in large samples for a long time, could continue to increase the complexity of the hypotheses they examine (path 5). Or, to come full circle, quantitative scholars could develop formal models to explain the empirical regularities they have documented so consistently, integrating their research with the growing body of formal theory (path 6). These paths would lead to new kinds of theory that are thick and integrated, integrated and general, or general and thick, any of which would be superior to the theories we have now. Even this much progress may be too much to hope for in the next generation of scholarship, but it would still leave us with the challenge of achieving all three criteria for good theory in one approach (paths 7, 8, and 9). This map of the difficult course ahead may be discouraging. Nevertheless, it makes it understandable why different approaches exist
now and defines some basic criteria for judging future progress in generating good theories.
Notes

1. John Gerring, in one of the most comprehensive discussions of “what is a good theory?” mentions
generality (“breadth”), integration (“analytical utility” or “logical economy”), and thickness (“depth”), but also
lists specification, accuracy, precision, parsimony, innovation, intelligibility, and relevance. John Gerring,

2. Gerring’s “specification” and “accuracy” are more pertinent to testing, discussed later in this chapter. I
discuss precision as a characteristic of conceptualization and measurement in another paper. Innovation and
intelligibility are lesser methodological virtues, and I do not consider parsimony a virtue at all.

3. Distinguishing between relevant and irrelevant cases is a crucial consideration. The rule of thumb is that a
theory should not be arbitrarily bounded. In comparative politics, a theory is arbitrarily bounded when its
application is restricted to certain objects for reasons that are not integral to the theory. A theory that applies
only to Asian countries for no potentially theorizable reason is arbitrarily bounded, but one that explains how
certain unique features of Asian countries condition the causal relationships is properly bounded.

Nicolson, 1953).

5. This does not mean that all of us must generalize, only that some of us must. Generalization is a collective
obligation.

6.“[The] ‘orthodox’ view of theories. . . has been called the ‘dual language’ conception of theories. In this
view there are two crucial components of a theory: a set of theoretical principles and a set of correspondence
rules which link the theoretical principles to observations, thereby providing an indirect or partial empirical
interpretation to the theoretical principles. A theory of this type is what is often called a

Greenstein and Nelson Polsby, eds., Political Science: Scope and Theory, vol. 1 of the Handbook of Political Science,
pp. 153-54.


9. All this presumes that the empirical referents of the propositions are clear. If is it is unclear how the
concepts in the theory correspond to observable phenomena, the theory is untestable for other reasons.


José Antonio Cheibub, and Fernando Limongi, "What Makes Democracies Endure?" Journal of Democracy 7:1


17. William of Occam’s instruction, “Pluralitas non est ponenda sine necessitas” (“Plurality should not be posited without necessity”), does imply that a more complex explanation would be preferred when it is necessary to account for the facts at hand.


30. Diamond, Linz, and Lipset, eds., *Democracy in Developing Countries*.


40. Collier and Collier, *Shaping the Political Arena*.


45. Przeworski and Limongi, “Modernization: Theories and Facts.” To be fair, the economic development hypothesis was originally embedded in the rather elaborate theory of modernization, which held that democratization was a trend parallel with development, urbanization, education, secularization, and industrialization. These trends were held to be causally linked in complex ways, although the nature of the actors driving the processes was not well specified. Lipset's seminal article developed various additional arguments about the relationship. Seymour Martin Lipset, "Some Social Requisites of Democracy: Economic Development and Political Legitimacy" *American Political Science Review* 53 (March 1959): 69-105. However, over the years, as the demands of testing thinned concepts, non-economic aspects of modernization encountered inconsistent confirmation, and the inevitability of modernization came into question, the theory was reduced to a simple, underspecified hypothesis.


