

Chapter 6: Formal Models and Theories

The approaches discussed in previous chapters--checklists, frameworks, case studies, and comparative history--can generate theory. However, they do so in a mostly inductive way, by generalizing from observations. Scholars working inductively stay close to what they observe whether they are defining concepts, developing theories, laying out a framework, generating hypotheses, or performing tests. The alternative is to work deductively, starting from fundamental axioms and using logic or mathematics to derive hypotheses (which can then be subjected to tests). Deduction and induction sound like opposites in principle, but in practice any attempt at theorizing must do some of both. Observation is impossible without a preexisting set of concepts, a rudimentary theory that tells one what to observe; and deductive theories are always inspired to some extent by real-world cases. Nevertheless, scholars differ greatly in the importance they attribute to the deductive and inductive stages of the theory-building process. Those who emphasize the deductive stages can be said to work with formal theories.

Formal theorists are sometimes critical of the more inductive approaches. In fact, many claim that only formal theories qualify as "theory" at all; they demote other general propositions to the lesser status of atheoretical "empirical generalizations" or "stylized facts." This is little more than intellectual snobbery that inflates the value of deduction and devalues the inductive processes required for building any good theory. However, formal theories have two characteristics that make them different from, if not better than, other sorts of theory. First, formal theories have clear and explicit assumptions. They have to; without premises, it would be impossible to derive implications. Stating assumptions explicitly makes it possible for other scholars to see the relationship between one study and another as a sort of tree diagram. The assumptions they share locates their origins on the same trunk or branch; the different assumptions they make push them out onto different

branches and twigs. Explicit assumptions therefore provide a way to achieve theoretical integration. Lacking this guidance, the more inductive approaches are harder to integrate. They end up with conclusions that are confined to provisionally bounded temporal-spatial domains that rarely coincide perfectly with the domains of other studies. Second, formal theories also have a heuristic: a hypothesis-generating procedure that is logically rigorous, or “truth preserving.” It allows the theorist to say with perfect confidence what *must* be true if the theory is true. It is invaluable for generating hypotheses for testing the theory.

Rational-choice theories—a subset of formal theories, but a dominant subset in among formal theories in comparative politics—have the additional characteristic of focusing of the intentions of actors. Defenders of rational-choice theory often argue that politics is an inherently goal-driven activity. Therefore, only theories that call attention to actors and their goals—the “microfoundations” of politics—can possibly capture the essence of politics. My own view is that any political theory that ignores intentional actors is merely incomplete, not entirely wrong, just as any political theory that ignores culture or institutions is incomplete. But it is true that no approach pays as much attention to actors and their intentions as rational choice theory.

Integration

The most distinctive methodological virtue of formal modeling is its emphasis on theoretical integration: making necessary connections among the propositions constituting a theory. We have already seen that checklists have no integration, frameworks are only partially integrated, and case studies and comparative histories are only very loosely and uncertainly integrated. In later chapters we will see that the large-sample, statistical approach rarely demands integration among the hypotheses it tests. However, whether formal modeling is actually an integrated approach depends on whether one is referring to one theorist’s model or the approach as a collective effort. The

propositions of each model are indeed well integrated, but only a few models are integrated with other models.

Of course, there are important benefits in the integration of the propositions in a single model. Formal models achieve internal integration by using game theory, logic, or other mathematics to derive predictions from their initial assumptions. A good example of game theory is Przeworski's analysis of strategizing by a moderate opposition and some reformers within an authoritarian regime (Przeworski 1991, 67-79). Przeworski assumes that each actor could ally only with an adjacent actor—the reformers with either the hardliners or the moderate opposition; and the moderate opposition with either the radical opposition or the reformers in the regime. The regime outcomes associated with each choice, and the associated payoffs for each actor, are as follows:

Figure 6.1: Przeworski's Extrication Game

		Moderates ally with:	
		Radical Opposition	Reformist Rulers
Reformist Rulers ally with:	Hardliners	existing authoritarianism 2, 1	authoritarianism with concessions 4, 2
	Moderate Opposition	democracy without guarantees to old regime 1, 4	democracy with guarantees 3, 3

Source: (Przeworski 1991), Table 2.1, p. 69.

Assuming full information (“everyone knows everything and everyone knows the same” (p. 62)) and an unrepeatable game without communication, Przeworski derives the conclusion that both actors would rationally choose the upper-right outcome—authoritarianism with concessions. The moderates anticipate that if they ally with the radicals, hoping for their highest payoff (4), the reformers would ally with the hardliners, improving *their* payoff from 1 to 2. However, the reformers anticipate that if they ally with the hardliners, hoping for the “authoritarianism with concessions” payoff (4), the

moderates will *not* prefer an alliance with the radicals, because this would reduce their payoff from 2 to 1. Of course, the moderates would be better off with democracy with guarantees (3), but as long as the reformers ally with the hardliners, this option is not available to the moderates. The moderates know that the reformers could make authoritarianism with concessions stick, and the reformers know that the moderates know that they could make it stick, so both actors settle on authoritarianism with concessions. Therefore, Przeworski concludes, there will be no transition under these conditions.

Acemoglu and Robinson (2001) is a good example of mathematical deduction. They write equations in which the incomes of the elite and the poor are functions of taxation, transfer payments, the degree of inequality, the growth rate, and other variables. Taxes represent a net gain for the poor because taxes fund transfer payments that are distributed equally to all, and for the same reason taxes are a net loss for the rich. When the poor govern (a “democracy”), they set the tax rate high; the elite could lead a coup to take over the government, but a coup is costly and risky. When the elite govern (a “nondemocracy”), they keep taxes low; the poor could lead a revolution, but a revolution is also costly and risky. The expected gains and losses from coups and revolutions enter into each actor's calculations. This allows Acemoglu and Robinson to predict that the poor will lead a democratic revolution against a nondemocracy when (I am simplifying their argument a bit) the economic benefits of expropriating the elite, raising taxes, and increasing transfer payments outweigh the expected costs of carrying out a revolution. The elite will lead a coup against the poor when the economic benefits of lowering taxes and preventing further expropriations outweigh the expected costs of a coup. The theory also links the marginal probability of transitions and breakdowns to observable conditions by making these costs and benefits depend on growth rates, degrees of inequality, and other economic factors.

There is a big difference, however, between the degree of integration achieved *within* models and the degree achieved *among* models. The approach is far less integrated than the models and theories that constitute it. Nevertheless, this approach is the democratization literature's best illustration of the process of constructing an integrated theory. An integrated theory begins with a set of simplifying assumptions that (to some scholars, at least) seem to be insights into some fundamental truths about politics. A group of scholars dedicates itself to exploring the implications of these assumptions for some political phenomenon of interest. The number of propositions entertained in this emerging approach grows in three different ways. First, some scholars elaborate pure theory: they progressively "relax" some of the overly simple assumptions and then rigorously derive the implications of this more complex set of axioms. This activity builds the "core" of the theory, and does so in a "truth-preserving" way that guarantees strict integration. The core theory is often still absurdly simplistic and usually not testable. Nevertheless, the pure theorists continue their efforts in the hope that eventually, when the core theory becomes sufficiently complex, it will become a reasonable approximation of reality.

Second, theory grows by modeling: attempts to build bridges between the core theory and reality. Modelers selectively interpret real-world cases so that the theory seems to apply to them, and also add to or revise propositions in the theory so that it does a better job of fitting the case or cases at hand. Modeling helps develop propositions in the "periphery" of the theory that tentatively seem plausible under certain circumstances. Modeling also lends credibility to the core theory by confirming the pure theorists' faith that the approach possesses the potential to approximate reality.

Third, theory cumulates as a result of systematic testing. Scholars derive testable hypotheses from their theories--things that must be true if the theory is true--and systematically gather and analyze evidence to see whether the hypotheses are consistent with the evidence or not. If the

evidence is consistent, it strengthens scholars' confidence in the theoretical propositions from which the hypothesis was derived. If the evidence is inconsistent, it weakens confidence in the theory and spurs a new round of theoretical innovation.

Have formal theories and models of democratization developed into a well-integrated body of theory? Beyond the shared traits that make them formal—clear definitions of the actors, their goals, the choices they face, and the use of deductive logic (Geddes 2003, 177)—formal theories of democratization are surprisingly disconnected. In other areas, such as theories of legislative behavior, formal theory may actually be well integrated, but for theories about the causes of democratization, integration is still a mostly unrealized goal. To be fair, formal theories of democracy are in their infancy: the first serious effort was published only in 1991. It is possible that one of the early efforts discussed in this chapter will blossom into a dominant theory that will subsume or marginalize all the others, leaving us with a well-integrated core theory. In fact, I believe that the theory developed in Carles Boix's *Democracy and Redistribution* (2003) is likely to become dominant within this approach. Nevertheless, it is too early to be certain of this.

There are two schools of formal theory about democratization.¹ Each makes a different assumption about the kinds of actors that struggle to define the regime and what their goals are. The first is what I will call the “positional” school. It assumes that the actors think about political regimes as ends in themselves. They usually rank the possible regime types in a preference ordering, and their preferences and position in the existing regime defines who they are—hardliners, softliners, opposition, and so forth. The second school I will call the “economic” school. It defines the actors according to their economic resources—usually a wealthy elite vs. poor non-elite—and assumes that the actors ultimately want to defend or improve their economic interests. Political regimes are assumed to affect economic interests in various ways, so actors prefer the regime that best serves

their economic interests. However, this regime preference is merely instrumental: a means to the actor's ultimate economic ends. Both approaches grew out of earlier, non-formal theories of democratization. The positional school formalizes and extends many of the arguments made first in the O'Donnell and Schmitter volume of the *Transitions from Authoritarian Rule* project (O'Donnell and Schmitter 1986). The economic school elaborates on Dahl's argument in *Polyarchy* that polyarchy arises when the costs of repression exceed the costs of toleration (Dahl 1971, 15-16).

The next sections of this chapter give an overview of the assumptions and conclusions of the most similar sub-schools within these two main schools. Within the positional school, I will discuss theories of liberalization and then theories of transition and/or survival, which I subdivide according to the number of actors involved. Following these long sections I summarize the economics school, which makes more uniform assumptions about the number of actors and the stage of democratization to be explained. Table 6.1 outlines the relationships among these sub- and sub-sub-theories and identifies the major works in each one. The chapter concludes with a critique of the benefits and limitations of formal theory for understanding democratization and some speculation about the future prospects for this approach.

Positional Models of Liberalization

Although formal theories in the positional school do not always explicitly acknowledge their debt to *Transitions from Authoritarian Rule* (O'Donnell and Schmitter 1986), the shared concepts, assumptions, and arguments make their intellectual lineage unmistakable. Like O'Donnell and Schmitter, almost all of these theorists assume that political liberalization must precede democratization; that liberalization becomes possible only when a split develops between the hardline rulers who favor continued nondemocratic rule and softline rulers who do not want democracy but favor a more open nondemocratic regime; that members of an opposition must

decide whether to collaborate with the rulers or protest their rule; and that the rulers must decide whether to open up the regime or repress the opposition. Also, in almost all these theories strategic choices must be made by both sides: the rulers' choice depends on whether repression would end the protests, and the opposition's decision to protest depends in part on what impact the protests would have on regime change. The difference between O'Donnell and Schmitter's analysis and the formal theories of liberalization is that O'Donnell and Schmitter opened the door to so many complex and dynamic possibilities that it became impossible for them to reach a definite conclusion. Their framework was wonderfully descriptive, even poetic at times, but it led them to conclude that transitions were full of uncertainty and inherently unpredictable. Formal theorists simplify more radically, permitting fewer possibilities, so that they can derive definite predictions.

Each theory focuses on one of two subgames, or both: a subgame between hardline and softline rulers and a subgame within the opposition. Typically, they then bring these analyses together in a meta-game to deduce what the rulers would do given the most likely choice of the opposition, and what the opposition would do given the most likely choice of the rulers. If rulers and opposition would agree on the same combination of opening or repression and cooperation or protest, then the meta-game has an equilibrium. The equilibrium is the expected outcome, according to these theories.

The theories of liberalization make different assumptions about the subgames that lead to different answers to a key question: Why would authoritarian rulers ever initiate a political opening? In the rulers' subgame, the simplest assumption is made by (Crescenzi 1999), who states that "Nature" chooses whether the regime is a hardline regime that will repress or a softline regime that will tolerate opposition protests. There is no real strategy involved in his subgame, so it does not derive an answer so much as assume one. (Przeworski 1991) has the softliners ("liberalizers") decide

whether to ally with the hardliners or open toward the opposition. In this strategic situation, Przeworski argues, the softliners would not lead an opening unless they knew that the opposition would accept their goal, a broadened dictatorship. Marks (1992), which is so faithful to O'Donnell and Schmitter's volume that it can be considered a formalization of the *Transitions* framework, models the bargaining between hardliners and softliners as a game of Chicken. Because disagreements between armed forces can be fatal, both factions would prefer mutual agreement to any disagreement. The prediction of this model is that if the hardliners are stronger or have more intense preferences, the softliners will go along with repression; but if the softliners are stronger or have more intense preferences, the hardliners will go along with toleration.

The central question of the opposition subgame is: Why would citizens ever protest during authoritarian rule, even in response to what seems to be an opening? Most of the theorists deal with this as a collective action problem: how can each citizen know whether enough other citizens will join the protest to ensure safety in numbers? Przeworski argued that citizens would *not* protest until some sort of signal occurred that increased their confidence that many others would protest and that protest would not be successfully repressed. He gave examples of such signals (the imminent death of a dictator, a looming economic crisis, or strong foreign pressures), but his theory left it up to the historical context to determine the kind of signal that would be effective (Przeworski 1986, 55). (Marks 1992) provided a more specific, although not more helpful, answer: the signal is a critical number of protesters. If protests (presumably by unusually daring citizens who are not acting rationally) reach a critical mass, it becomes rational for all citizens to join them. However, Marks also argued that if a higher critical threshold were reached, the protests would provoke a backlash from the regime, in which case it would be rational for everyone to stay at home. Not knowing what these critical thresholds are, it is hard for us to make a prediction. Lohmann (1994), analyzing

turnout in East German protests in 1953 and 1989-90, provided a more helpful answer. She recast the opposition subgame as a signaling game in which every citizen was both a sender and a receiver of messages about the depth of discontent with the regime. She argued that the signal is not aggregate turnout, but the gap between actual and expected turnout and the kind of citizens who turn out. Citizens discount protests that appear to be organized or that involve people who are likely to protest with little provocation. They take more seriously any protest by those who are less likely to protest, especially when it seems to happen spontaneously. Early spontaneous turnout of moderates sends a strong signal that discontent is widespread, and this encourages others to join.

In the meta-game, Przeworski (1991) laid out an extended-form (sequential) game in which (1) the softliners liberalize (or not), (2) the opposition cooperates with broadened dictatorship or mobilizes in protest, and if the opposition mobilizes, then (3) the softliners either repress (successfully or not) or become pro-democratic reformers. He argued that because the softliners would not sponsor an opening if it was likely to lead to democracy, a transition was possible only if the softliners miscalculated (about the opposition's willingness to acquiesce in a broadened dictatorship or about the rulers' chances of repressing successfully) or secretly were, or subsequently became, sincere advocates of full democracy (in which case they were not, or ceased to be, liberalizers).² Marks (1992) reached a more optimistic conclusion than Przeworski's in which liberalization can happen without an accident or a change of heart. First, he found that there are some non-strategic situations in which the rulers would always tolerate or always repress, regardless of how the opposition behaves. Second, he identified three strategic situations in which the rulers would eventually decide to tolerate rather than repress. In one of these situations, the opposition's determination to resist repression leads the ruler to tolerate. In the second situation, the ruler's willingness to answer protest with repression persuades the opposition to acquiesce to a liberalized

dictatorship. The third situation is a Prisoner's Dilemma that would normally lead to violence. However, if both actors foresee a long struggle and value liberalization much more than violence, they can both compromise on a liberalized dictatorship.

Theories of liberalization provide answers that are interesting but not entirely satisfying. For example, why does a split develop between hardliners and softliners? Crescenzi assumes an answer, Przeworski all but proves that any liberalization would contradict his model, and (Zielinski 1995) appeals to a *deus ex machina* (the Soviet Union) that could not be part of any general theory. Marks's chicken game is the most sophisticated answer, but even this answer starts from the assumption that hardliners and softliners already exist and have different goals; so the origins and the nature of the split are assumed, not explained. And what makes it rational for citizens to protest against a dictatorship? The notion of a signal is intuitively appealing, but Przeworski's ideas about signaling are vague, Marks's are circular, and Lohmann's cannot be applied beyond the now-extinct East Germany without additional interpretive effort. However, it would be too harsh to dismiss these efforts. Every theory is incomplete in some way, and every deductive theory has to start with assumptions of some kind. These theories tell us when a hardliner-softliner split would lead to liberalization and when it would lead to more repression. The job of explaining how the split arises can be left to other works and other scholars. This "upstream integration" has not yet been done for theories of liberalization. However, as the next section shows, there has been "downstream integration" linking theories of liberalization to theories of transition and survival.

Positional Theories of Transition and Survival

In the positional school, the existing theories of transition and survival appear to be integrated in an upstream-downstream hierarchy. Farthest upstream are the simple theories with just two actors—a ruler and the opposition. Downstream there are more complex theories with a ruler

facing two opposition sub-actors, some with a single opposition facing two ruling sub-actors (the familiar hardliners and softliners), and a couple of theories with ruling sub-actors facing opposition sub-actors. The logic of the hierarchy, then, is that the downstream theories relax the assumption that one of the upstream actors is unitary. This allows the downstream theory to explore a subgame within the opposition or ruling camp, bringing each downstream model progressively closer to the complexity of real-world politics. However, this hierarchy is interrupted. The links among theories are broken by the idiosyncratic assumptions made in each theory that remove it from the most direct line of descent. The genealogy linking these formal models lacks direct parent-child relationships; instead, it is a spotty genealogy containing only great-aunts, third cousins, and other distant relations.

The basic sequential game is the same in the two-actor models (Sutter 1995, ; Swaminathan 1999). A unitary opposition either challenges the ruler or acquiesces in nondemocratic rule. If the opposition challenges, the ruler can either surrender or repress. If the ruler represses, there is some probability (a “lottery,” in formal theory jargon) that the opposition will win and a transition will happen; otherwise, nondemocratic rule continues.³

At this point each model adds wrinkles that make it unique. Swaminathan (1999) assumes that the power of the ruler is constantly declining and that of the opposition is constantly increasing. Given this assumption, a transition is inevitable. The only questions are when it will happen and whether it will be peaceful or not. If a risk-acceptant ruler faces a risk-acceptant opposition, then the opposition will challenge, the ruler will repress, and one actor will win the conflict. This is the sole conflictual path to democracy. Even if the opposition loses at first, it will win a later conflict because of its growing power and the ruler’s growing weakness. Peaceful (“negotiated”) transitions will happen when both actors are risk-averse or when one is risk-averse and the other is risk-acceptant

(subject also to the costs of conflict, the actors' discount rates, and other assumptions). The only application of this model is to explain the timing of a transition and the likelihood of conflict along the way. In Sutter's (1999) model, the opposition can choose to negotiate rather than challenge or acquiesce, and the crucial consideration is not risk aversion but how generous a deal the opposition can offer the ruler in exchange for abdication. If the ruler would prefer such a deal to losing a costly struggle to stay in power, he will negotiate the terms of his abdication. The possibility of the ruler retaking power after abdication helps ensure that the opposition will honor the terms of the deal. Sutter also introduces a third actor, a foreign power, which can offer the ruler asylum in case of abdication. He finds that the possibility of asylum would improve the chances of a transition if the ruler would accept it. However, the foreign power has little reason not to renege on the offer, so the dictator is less likely to accept, which in effect narrows the opposition's choices to negotiating or fighting. The conclusions of these models hinge crucially on the extra assumptions they make, which prevent us from integrating them into a single theory. Without deriving new models, we cannot say whether Sutter's conclusions depend on the actors' tolerance of risk or how Swaminathan's ruler would respond to the prospects of going into exile or getting a share of the former opposition's spoils of office. This is less integration than one would expect from formal theory.

Moving downstream, we encounter some three-actor games involving the opposition and two ruling sub-actors: the hardliners and softliners.⁴ The crucial element in both games is how various actors cope with uncertainty about how other actors are likely to respond. Crescenzi (1999) focuses on the opposition's uncertainty about whether it is dealing with a hardline or softline regime (which is determined by Nature, as mentioned above). The opposition must guess whether hardliners or softliners are in control based on any liberalization that is adopted. Without liberalization, the opposition will conclude that the rulers are hardliners and will acquiesce.

However, a pacted transition can occur if the rulers liberalize, the opposition negotiates with it, and the rulers turn out to be softliners. A revolutionary transition can occur if the rulers liberalize, the opposition tries to negotiate but is rebuffed, and then the opposition radicalizes and (if it is strong enough) wins a violent struggle. (Zielinski 1999) models a negotiation between the softliners and the opposition over the degree of reform, on a continuum ranging from the status quo through liberalization to transition. In his model, it is the softliners who are uncertain: in this case, about whether the hardliners will ratify a deal struck with the opposition or intervene to reimpose a dictatorship. As long as the softliners accurately gauge the limits of the hardliners' tolerance, they can use the threat of a hardliner intervention to induce the opposition to support a softliner degree of reform. But if the softliners misjudge the hardliners' intentions, their reform agreement will be overturned and dictatorship will continue.⁵

A different path downstream leads to models that retain a unitary ruler but subdivide the opposition. However, these authors subdivide the opposition in different ways. Casper and Taylor (1996) have the rulers (the "defender") face off against a "challenger" and the "mass public" while Weingast (1997) pits the ruler (the "sovereign") against two undifferentiated groups of citizens, A and B. Because neither group cannot be equated with Casper and Taylor's challenger or their mass public, these two theories are not really integrated with each other; all they have in common is the assumption that the ruler is unitary and the opposition is not.

In Casper and Taylor's model, there is a familiar process in which the opposition either challenges or backs down and the rulers either give in or repress.⁶ The chief difference of this model is that within the opposition there is a mass public that mediates the contest between the rulers and the challenger. The more the mass public signals its support for democracy, the greater the bargaining leverage the challenger can wield. Based on this mass public reaction, the rulers and

challenger update their perceptions of their strength and a new round of negotiation begins. This stage ends when either rulers and challenger agree or one of them ends negotiations in order to pursue a violent solution. Transition becomes unlikely when the rulers repress early on, silencing the mass public's cues. Transition is possible when strengths are initially evenly matched. The rulers recognize early on that they cannot stop the mass public from supporting the challenger, so they compromise, negotiating rules to protect themselves after the transition. These are successful transitions, but according to Casper and Taylor they have trouble with consolidation. Transition is also possible when there are initially diverging preferences that make the rulers overconfident and the challenger uncertain. But the mass public consistently supports the challenger, so perceptions of strength shift. Eventually the rulers lose their ability to impose protective rules, so the transition is more complete and consolidation is more successful.

In Weingast's (1997) model, the rulers (the "Sovereign") deal with two groups of citizens (A and B) of relatively equal status—unlike the unequal challenger and mass public in Casper and Taylor's model.⁷ Nondemocratic rule lasts as long as the rulers can divide and conquer the citizens, coopting a large enough group to prevent a successful revolt by the other. Transition (to limited government, not necessarily democracy) becomes possible only when A and B coordinate their resistance to the rulers' transgressions against citizens. Limited government is consolidated when A and B find a self-enforcing solution to their coordination dilemma. The self-enforcing solution is a pact that commits both groups to resist any transgression by the rulers against any citizen. This pact is self-enforcing because either group can credibly threaten to help the rulers punish the other group if that group shirks its responsibilities under the pact. Again, these models are cousins (distinct in more than one respect) rather than siblings, so we cannot say for certain why their conclusions differ. Nevertheless, it seems likely that the coordination dilemma that Weingast considered so

central is an artifact of his assumption that the two opposition actors are of equal status and act simultaneously. Casper and Taylor's more hierarchical and sequential model does not pose this dilemma. Readers must decide for themselves which set of assumptions is more realistic.

Finally, another set of theorists works a bit farther downstream, where both rulers and opposition are subdivided. In their theories there are more actors, but they usually interact in pairs. The outcome of the game is determined by *which* pair of actors is involved. These theories do not tell us how some of the possible pairs are selected or how to know which pair is relevant in a specific case; this is left up to the skill of the analyst who wishes to apply these models. These theories are therefore less complex and less complete than they appear.

The seminal model of this type comes from a later section of Przeworski's *Democracy and the Market*. The essence of his model is that there are four actors, arrayed on a democracy-dictatorship continuum as follows. At the most democratic pole are the radical opposition (which Przeworski calls the "radicals"), followed by the moderate opposition ("moderates"), then the softliners ("reformers"), and finally the hardliners. Przeworski makes the moderate opposition and the softliners the pivotal players: each chooses which adjacent player to ally with, and their choices determine the outcome. Four outcomes are possible. 1) If the softliners stick with the hardliners and the moderate opposition remains allied with the radical opposition, then the old authoritarian regime survives unchanged. 2) If the softliners stick with the hardliners but the moderate opposition tries to ally with the softliners, then the softliners are able to persuade the hardliners to accept a liberalized authoritarian regime. 3) If the moderate opposition tries to ally with the softliners and the softliners reciprocate, then there is a transition to a democratic regime that includes some immunities for the outgoing authoritarian rulers and each actor is able to persuade its more extreme counterpart to accept this deal. 4) If the softliners wish to ally with the moderate opposition but the

moderate opposition sticks with the radical opposition, then there is a democratic regime without guarantees for the outgoing rulers.

The two pivotal actors choose their alliances strategically, in a one-shot game (Przeworski 1986, 72-73). The moderate opposition knows that if it allies with the radical opposition, the softliners would always be better off sticking with the hardliners, resulting in the opposition's worst-case scenario: survival of the authoritarian regime. Therefore, the moderate opposition takes the risk-averse course of seeking an alliance with the softliners. The payoffs to the softliners, however, depend on whether they remain united with the hardliners or can act independently. If they are united, then the softliners stay with the hardliners, producing authoritarianism with concessions. It is only when there is a division between hardliners and softliners that the softliners are better off allying with the moderate opposition, leading to democracy with immunities.⁸

Although this model has been very influential, it is not the best example of formal theorizing, because some of its assumptions are implicit and some of its logic is not fully developed. Przeworski does a good job of defending his assumption that this is a one-shot game, but it is not clear why the reformers must act before the moderates rather than the other way around or having both actors choose simultaneously. Also, it may be unnecessarily restrictive to assume that the two centrist players are pivotal and that they can ally only with adjacent actors. Indeed, at one point Przeworski equivocates on this by speculating about how the outcomes could differ if the moderate opposition allied with the hardliners or the softliners allied with the radical opposition (Przeworski 1991, 73-79). In a way, it becomes a three-player game when the softliners' payoffs depend on whether there is unity with the hardliners or not. But Przeworski never tells us what would happen if the radical opposition spurned an alliance proffered by the moderate opposition. Finally, what is least clear and most crucial is the association of regime outcomes with the actors' choices. Would

democracy always result if the moderate actors allied? Would united rulers always defeat a united opposition? We just have to take Przeworski's word for it.

Josep Colomer (1991 and 2000) builds a theory around six actors, obtained by subdividing the hardliners and softliners into moderate and radical factions. As in Przeworski's model, only two or three actors participate in any one game and the outcome of the game is determined by which actors are included. However, unlike Przeworski, Colomer makes explicit assumptions about the connections between regime outcomes and actors' choices. His actors are defined by their preference ordering over the same three transition options: democratic rupture (D), intermediate reform (I), or nondemocratic continuity (N). ("Intermediate reform" could be either Przeworski's authoritarianism with concessions or his democracy with guarantees.) Each actor also has a unique ordering of preferences over all the "collective outcomes" that would result from the choices made by pairs of actors: DD, DI, II, DN, IN, and NN. Colomer uses these preference orderings to assign payoffs to each actor, and these payoffs allow him to solve for the equilibria (if any) of all the possible games. When both actors agree (DD, II, or NN), the outcome is stable--democracy, an intermediate regime, or nondemocracy, respectively. When the actors disagree (DI, DN, or IN), the outcome is conflict.

What is distinctive about Colomer's theory is its open-endedness. The 15 two-player and three three-player games that he discusses can lead to any of the possible outcomes, depending on which actors are in the game. If there were a reliable way to know which actors are relevant in any given situation, this would be a very useful (or at least testable) theory. But the theory does not tell us how to know which actors are relevant. Also, the assumption that only two or three actors of the six possible ones are ever relevant in any given situation strains credulity.

All of these regime preference theories share an additional problem with the O'Donnell-

Schmitter-Whitehead framework that originally inspired them: the near-circularity of defining actors in terms of their regime preferences. As discussed in Chapter 4, it is difficult to identify these actors in actual situations if they are not defined in terms of their institutional position or social class.⁹ Furthermore, if actors are allowed to change their preferences during the process (as both (Przeworski 1991, 73) and (Colomer 2000, 69) contemplate in some passages), the theory loses all explanatory traction. The next school of formal theorizing avoids this problem.

Economic Theories of Transition and Survival

With one small exception, formal theories of democratization that assume that actors value regimes only as a means to economic ends are better integrated. The family relationships are easier to trace because they share so many more assumptions in common. Almost all assume that there are just two actors, an elite and the poor; all define democracy as a regime in which the poor make economic policy and non-democracy as a regime in which the elite make policy; all deal with policies such as tax rates and distribution of the economic surplus; and all propose that actors choose the regime that maximizes their economic returns. These theories sometimes arrive at opposite conclusions due to other assumptions that are different, but the resemblances are so strong that any reader of these articles will encounter much that is familiar.

The one small exception is Youssef Cohen's game-theoretic explanation of the breakdown of democracy in Brazil (1964) and Chile (1973) (Cohen 1994). Cohen defines the actors as the extreme left, the moderate left, the moderate right, and the extreme right. He argues that the moderate left faces a prisoner's dilemma when deciding whether to break with the extreme left, and the moderates and extremists on the right face the same dilemma. One would think that the two moderates would choose to cooperate in supporting a program of moderate economic reform. Unfortunately, Cohen argues, the moderate left fears that the moderate right would betray it and the

moderate right fears that the moderate left would betray it, leading to their least-preferred outcomes. Therefore, both prefer solidarity with their respective extremists. This results in polarization and deadlock, which Cohen equates with the breakdown of the democratic regime. The literature on the prisoner's dilemma has found that such outcomes can be avoided if the actors expect to play the game indefinitely and do not discount the future (Fudenberg and Maskin 1986). Cohen, however, like Przeworski, argues that regime transition decisions are non-repeatable one-shot games.

The other economic theories that developed after Cohen's book start from scratch, with very different assumptions. Whether they acknowledge their pedigree or not, all develop Dahl's 1971 argument that transitions occur when the costs of repression exceed the costs of toleration (Dahl 1971, 15-17). The first formalization of this model was by Feng and Zak (1999), who translated the costs of repression as spending on the police, and the costs of toleration as the tax rate. However, they complicated the model much more than this. The ruler ("autocrat") could not only vary the tax rate and spending on the police; he could also vary levels of spending on education. For its part, the opposition ("agents") could decide how many hours to work productively and how many hours to spend demonstrating against the ruler. Ultimately, agents seek to maximize their lifetime consumption. Because demonstrations destroy assets, they shrink the tax base and undermine the ruler's ability to pay for police. Feng and Zak argue that a transition happens when "antigovernment demonstrations overwhelm the autocrat's ability to maintain public order" (169). After some complex derivations, they conclude that 1) a transition should happen immediately if most agents are rich; 2) quickly if the economy grows and there is a large number of agents just below the threshold at which it become profitable to demonstrate; or 3) slowly if there is such great inequality that most people are far below the threshold and the economy is not growing; 4) transitions are also more likely when the government invests in education (which boosts productivity and therefore

assets) or police (because it prevents demonstrations) and 5) when people have a "preference for democracy."¹⁰ The theory is very simple, but the fact that most of its predictions coincide with the known empirical relationships is encouraging.

Rosendorff (2001) made three modifications to Feng and Zak's model. First, whoever is ruling (the elite in an autocracy, the median worker in a democracy) distributes lump-sum shares of collected taxes to all, equally. (As in most of these models, democracy--rule by many--necessarily implies a higher tax rate.¹¹) Second, if the elite loses control, they will be expropriated entirely and the proceeds will be distributed equally.¹² These modifications provide extra incentives for workers to demonstrate and for the elite to spend more on security. Rosendorff's conclusions, like Feng and Zak's, focus on the impact of inequality and economic growth. He finds that as inequality worsens, both sides spend more on protecting or contesting the autocracy. But if equality increases beyond a certain threshold, then the elite's costs of maintaining the regime exceed their expected losses from redistribution, so they agree to a transition. So transitions are likely when the income distribution becomes flatter, the growth rate is low, or the size of the workforce decreases.

Acemoglu and Robinson (2001) published a still more elaborate version of this basic model. One major difference in their assumptions, however, is that both actors rely on carrots more than sticks. In their model, an autocratic elite government does not normally prevent a democratic revolution by investing in repression but by redistributing tax receipts more generously, and a democratic government prevents an authoritarian coup by lowering the tax rate enough to pacify the wealthy elite. Various economic conditions impinge on the probability of democratic transitions. Other things being equal, transitions to democracy are more likely when a severe recession denies the elite the resources it needs to prevent a revolution, and also when inequality is high. Transitions are less likely when the economy is growing fast or when the elite invests in repression, especially

when inequality is high. Once established, a democratic regime is more likely to survive when the ruling poor keep taxes low; they may even constitutionally limit the tax rate to reassure the elite. Such reassurance is especially welcome when recessions are frequent. Democratic stability is also favored when the economy depends on investments that pay off only as long as democracy survives. Acemoglu and Robinson also propose that a relatively equal distribution of assets would enhance democracy's survival prospects, although they are aware that the threat of massive redistribution to attain this level of equality could well provoke an elite coup. Curiously, one of their conclusions is that there should be no association between average wealth and democracy unless wealth is associated with economic volatility or the power of the poor.

Gould and Maggio (2003) develop a similar, but simpler, model that proposes an explanation for an empirical finding of Przeworski et al. (2001): wealth does more to help countries *remain* democratic than it does to help them *become* democratic. They assume, in the spirit of the works just discussed, that 1) in a dictatorship, elites maximize earnings but risk a costly revolution; 2) non-elites would earn more in a democracy, but a democratic revolution would be costly to them; 3) in a democracy, elites earn less but face no risk of regime change; and 4) non-elites in a democracy earn more but face the risk of a coup. Using Kahneman and Tversky's prospect theory, Gould and Maggio argue that both actors are loss-averse (Kahneman and Tversky 1979). That is, elites would tolerate democracy to avoid the losses associated with an unsuccessful coup, and non-elites would take the risk of opposing a coup to avoid the losses they would suffer under dictatorship. This theory predicts that democracy is "stickier" as countries become wealthier because the loss aversion is more pronounced when the stakes are higher.

Zak and Feng (2003), building explicitly on their own 1999 article, Acemoglu and Robinson (2001), Lohmann (1993 and 1994), and several other contributions, generate even more plausible

predictions by emphasizing the distinction between income and assets. As in their 1999 article, citizens choose a mix of productive work and destructive demonstration that will maximize their lifetime consumption, and an autocrat chooses a mix of taxation, investment in public goods, and paying for police that will maximize the economic growth rate. And as in their 1999 model, transitions occur when the autocrat cannot raise police spending enough to prevent demonstrations.

These assumptions are still a gross simplification of reality, and some may even be false. Nevertheless, the predictions that flow from them mirror empirical relationships faithfully in a surprising range of respects. Zak and Feng's central distinction is between a "wealth effect"--wealthier citizens demanding civil liberties when the economy is growing--and an "income effect"--middle-class citizens protesting low incomes resulting from economic contraction. The authors find that the citizens who will spend more time on demonstrations are those with large assets but low wages, such as the middle class and students. Transitions happen when either the economy is growing so rapidly that the wealth effect dominates the income effect, or there is such a deep recession that the income effect dominates the wealth effect. In either situation, demonstrations outstrip the police protection that the autocrat can afford. In contrast, slow, positive growth retards transitions. Zak and Feng also find that unequal wages inhibit transitions but unequal wealth hastens transitions, provided that the autocrat spends too little to prevent his overthrow.

Already there is a noticeable trend: the early theories in this school were overly simplistic, but the more complex ones begin to approximate recognizable tendencies in the real world. The economics theory that approximates reality in the widest range of respects is that of Carles Boix (2003), which not only brings together the best insights of its rational-choice predecessors; it also formalizes several ideas from comparative historical research on democratization. This theoretical synthesis is a very impressive achievement. His efforts to test hypotheses from the model (which will

be discussed in chapter 11) make it even more impressive.

Like other economics theorists, Boix assumes that the wealthy rule in an authoritarian regime and the poor in a democracy, that governments make decisions about taxes and spending on repression, that taxes are redistributive, that inequality and growth rates affect these decisions, that violence is costly, that elite assets are confiscated after a revolution, that assets pay dividends, and that actors prefer the regime that best serves their economics. However, Boix makes several assumptions that are distinct from those made by other theorists in the economics school. The key one is that assets vary in their specificity: some assets, such as agricultural land or oil fields, are specific to a country while others, such as labor, technology, and commercial goods, are mobile across national borders. This distinction is highly consequential, as we shall see. Boix also adopts the important distinction between wealth and assets, which only Acemoglu and Robinson (2001) and Zak and Feng (2003) also make.¹³ This useful assumption also helps distinguish between democratic and communist regimes, which most other works in this school fail to do. A third distinct assumption is informational asymmetry: the wealthy know how much repression costs them, but the poor do not; yet the poor are better informed about how well organized they are, and therefore about their ability to revolt successfully. Boix argues that sudden shifts in the distribution of information about the chances of successful repression or revolt explain the timing of regime changes. Finally, Boix is the lone economics theorist to relax the assumption that there are only two classes. After developing a basic two-class theory, he adds a middle class, which introduces two more possible regime outcomes (limited democracy and a milder revolutionary regime, in addition to democracy, right-wing authoritarianism, communism, and civil war) and opens up the possibility of class alliances.

The variety of predictions that Boix derives from his theory, and their correspondences with

at least some empirical cases, make this model so promising that it has the best chance to become the dominant rational-choice approach to explaining democratization. He predicts that capital mobility forces even democratic governments to keep taxes low, which helps the wealthy make peace with democracy. This is why, he claims, economic development favors democracy: not so much that incomes rise, but that wealth shifts from highly specific investments in land to more mobile investments in industry, commerce, finance, and technology.¹⁴ This reasoning also eliminates the paradox of premodern democracy in societies with relatively low levels of development, such as ancient Athens and the nineteenth-century United States (Boix 2003, 41-44). It also accounts for the lack of democratization in wealthy oil-exporting countries, which have highly specific assets. Economic development can also favor democracy if it reduces inequality, because the more equal the society, the less a democratic government would need to raise taxes in order to redistribute wealth, and the less the wealthy would prefer to install an authoritarian regime. Boix argues that European transitions coincided roughly with periods of growing equality in the United Kingdom, Denmark, Norway, and Spain (Boix 2003, 36-38). Boix's assumptions about the impact of working-class organization on the costs of repression help explain the timing of Western European transitions in the first decades of the twentieth century, when unions and leftist parties began to form (Boix 2003, 52-53). Also, his emphasis on asset specificity allows him to model situations in which large landowners would oppose democracy while the urban bourgeoisie would favor it, much as Barrington Moore proposed (Boix 2003, 53-56).

Whether these correspondences between theory and facts are systematically true or merely selective examples will be discussed in chapter 7. The most relevant conclusion for the present chapter, however, is that Boix's model is the best illustration of the distinctive potential of formal theory to achieve theoretical integration, which is a necessary characteristic of good theorizing.

Evaluating Formal Theories

Although formal theories have been part of political science since the 1950s, they have been applied to democratization only since the early 1990s. It would be unfair to judge an entire approach based solely on the record of one of its youngest applications, so I base my evaluations on tendencies that are typical of formal theorizing on topics that have a longer track record, and I do this in a spirit of charity and hopefulness. In brief, formal theory does a relatively good job of producing integrated theory, but a relatively poor job of producing thick theory that is empirically general. I emphasize "relatively" in this conclusion because formal models of democratization are neither well integrated nor completely devoid of thickness or general confirmation; they are merely better integrated, thinner, and less well tested than other approaches. Furthermore, within the formal approach some of the economic models are better integrated, thicker, and better tested than the positional models.

Thickness

Formal theories are sometimes criticized for oversimplifying reality. To be fair, we must be clear: all models simplify reality. Models select certain features of reality to highlight and ignore other features (Morton 1999, 39). A reasonable simplification is one that calls attention to the most important features of a process while ignoring less important features. However, some models simplify to such a degree that they miss the most important parts of the story, while others try to pass off contradictions of the real story as simplifications. It is reasonable to say that Al Gore lost the 2000 presidential election because it was so close to a tie that the imprecision of Florida's balloting system created legal challenges that allowed the U.S. Supreme Court to decide the election. To say that he lost because of the butterfly ballot or the Nader candidacy is an oversimplification. To say that he lost because he wanted George W. Bush to win is simply false.

The concepts and propositions of formal theories of democratization are extraordinarily thin. Some positional models refer to the "rulers" (Swaminathan 1999, ; Zielinski 1995) or "sovereign" (Weingast 1997), the "opposition" (Marks 1992, Sutter 1995), or the "mass public" (Casper and Taylor 1996) as monolithic, undifferentiated blocs, ignoring differences between hardliners and softliners, unions and parties, left and right, that are likely to be consequential. Of course, some other positional models break down these actors a bit: Colomer (1991, 2000) goes the farthest, theorizing about six actors. But even in the case of Colomer, the six actors are defined purely by their rank-ordering of just three permissible "strategic preferences for a greater or lesser degree of discontinuity with the existing regime" (Colomer 2000, 33)--democratic rupture, intermediate reform, or nondemocratic continuity. This tells us nothing about the actors' positions on an economic left-right continuum, their popular support, the charisma of their leaders, their internal unity, or how well organized they are. Defining actors solely in terms of regime preferences amounts to an assertion that these other qualities do not matter for democratization.

Among economic models, the most disturbing thin concept is democracy itself. Most of these models define democracy as a regime in which the poor (or the median voter, who is poor) make economic policy, including setting tax rates and redistributing assets. Acemoglu and Robinson (2001, 941-2) do not distinguish between a democratic transition (which they call a "revolution") and a utopian communist revolution: "After a revolution, poor agents expropriate an additional fraction. . . of the asset stock of the economy. . . . We also assume that the rich lose everything after a revolution. . . ." This bears almost no resemblance to the regimes that we refer to as democracies. Boix comes closer to reality by distinguishing between communism, in which "the poor rule after expropriating all the wealthy's capital," and democracy, in which "property is preserved and everybody votes on the tax rate" (Boix 2003, 23). However, this definition of democracy fails to

mention civil liberties, political rights, representative institutions, or any other procedures; it only alludes vaguely to some kind of election. Boix reduces democracy to the number of economic classes that participate in setting the tax rate.

These thin concepts become the building blocks of thin theories. The positional models are thin in the sense that they limit their attention to the final moment of a transition, when everything but the outcome has been decided. The actors have been selected, the choices available to them are identified, their preference orderings are defined, their time horizons and access to information are known; all they have to do is strategize and reach a decision. Given these highly structured situations, these models can make a prediction. But these models do not tell us how these actors, out of all the many possible actors in a society, came to be the ones making the final decision; why the choices offered them are the only ones available; why they prefer some outcomes to others; or other important details about how the game is structured. A thick theory would address these other issues.

In some ways, the economic models are thicker. Unlike the positional models, which tend to take shape as simple 2 X 2 strategic-form games (Check Krebs for proper terminology) with discrete choices and ordinal payoffs, the economic models are written as algebraic utility functions for each actor. These functions allow the payoffs to vary continuously in response to tax rates, economic growth, the rate of return on investment, and other variables. Furthermore, each player's utility can be calculated at any point in time, not just at the moment of decision. So these models are structured less rigidly, and are therefore less narrow. The better models also approach thickness by taking into consideration a richer variety of economic variables. Zak and Feng (2003), for example, incorporate into their model wages, assets, returns on investment, the rate of economic growth, tax rates, tax receipts, redistribution, spending on education, spending on police, the effectiveness of spending on police, productivity, time devoted to productive work, time devoted to destructive protests, wage

inequality, and asset inequality. Boix (2003) adds to this mix capital mobility, which enables him to distinguish among economic sectors within the wealthy class; unionization; and cross-class coalitions. Nevertheless, these models are still thin in their exclusive focus on economics. They ignore traditions, institutions, non-economic values, demonstration effects, and non-economic power resources such as moral authority (though Boix--alone in this school--also considers mass organization and access to information). Because of this theoretical thinness, the positional and economic models of democratization are incompatible with each other. Positional models claim that actors value regimes in themselves while economic models assert that actors value regimes only as means to economic ends. Only one of these can be a reasonable simplification; the other must logically be an oversimplification. Combining economic and non-economic motivations would be a move toward thickness, but so far formal models have avoided assuming that actors have conflicting motives.

Thickening these thin theories would make them more realistic and change their premises from oversimplifications or contradictions of fact into reasonable simplifications. However, it is too much to expect any single model to approach a realistic degree of complexity all by itself; it must build on simpler and therefore less realistic models worked out by others--the kind we have now. There is hope that eventually formal models will reach this degree of sophistication. But simply increasing the number of models is not enough; rather, they must be *integrated* so that each model improves on the ones that came before it.

Integration

If we treat all formal theorizing about democratization as a single theory, the purely theoretical core is very small--only the assumption that actors are rational. There are no universally shared assumptions about how many rational actors there are, what their goals are, what choices

they face, or even what kind of rationality they practice. However, if we consider the positional models and economic theories separate theories, their cores become a bit more elaborate. The positional models assume that actors are defined by their regime preferences, that some actors are in the regime and some in the opposition, that a regime change involves a loss of power for the rulers and a gain in power for at least some in the opposition, and that liberalization precedes democratization. Beyond this, all those modeling liberalization share some additional assumptions, and all those modeling transitions or survival share some other additional assumptions. However, few theorists in this school agree on the number of actors, the alliances available to them, the choices available to them, or which game they are playing. Nevertheless, despite the different assumptions, the positional models tend to conclude that 1) the outcome of the game depends heavily on the regime preference orderings of the people who happen to be leading the regime and the opposition in a given country; 2) hardliners will not liberalize if they know it will lead to democracy; 3) some kind of signal is necessary to embolden the opposition to challenge the regime; 4) if the odds of success shift in favor of the opposition, a transition eventually becomes inevitable; and 5) the actor with better information has a better chance of imposing its preferred regime. This may not be perfect integration, but it is enough to constitute an interesting school of thought.

In the economic school there is an even larger body of core theory, already described above. Still, it is not complete integration: each article adds wrinkles that others ignore. Only Zak and Feng (1999 and 2003) suppose that protesters destroy productive assets; only Acemoglu and Robinson (2001) consider whether commitments to lower taxes are credible; only Gould and Maggio (2003) use prospect theory; only Rosendorff (2001) and Boix (2003) take capital mobility into account; and Boix never cites Zak or Feng. Despite these different assumptions, the economic theories tend to agree that 1) democracy is more likely to be born and survive in egalitarian societies,

2) democracy is more likely to survive when elites can shift their assets abroad, and 3) regime change is a rare event because challenging a regime is costly and risky.

Generality

Formal theories may seem to be general because most are expressed in universalistic terms. However, this universalism is almost entirely an aspiration rather than a reality. This approach employs concepts defined at a high level of generality, such as "softliner" and "the poor," and (aside from the models of particular cases) they tend not to circumscribe their domain of applicability. The work with the most extreme generality of aspiration, Barry Weingast's article on democracy and the rule of law, claims to shed light not only on the English Glorious Revolution of 1689, but also on social capital in regions of Italy, consociationalism in divided societies, the Salvadoran peace accords, and the Missouri Compromise of 1820 (Weingast 1997). But in order to be truly general, these theories must not only be cast in general terms; they must be *empirically* general. That is, they must be consistent with all the known evidence in their domain of applicability and not restrict their domain arbitrarily. Few formal models or theories even come close to meeting this standard.

Working toward empirical generality requires testing. However, there is disagreement about what should be tested. It is common for defenders of rational-choice theory to urge others to test the predictions but not to test the assumptions. The first line of defense for this advice is the claim that many of the assumptions of formal theories are unobservable, such as the subjective value of payoffs and the strategic reasoning of the actors. This would be an air-tight defense if not for the fact that some assumptions of these theories are in fact observable. We can know how many actors are parties to negotiations over regime change, whether it is realistic to say that the median voter sets tax policy, whether authoritarian regimes have lower tax rates than democracies, and so on. When these assumptions can be confronted with evidence, they should be. The second line of defense for

not testing assumptions is that the worth of a theory is better gauged by the accuracy of its predictions than by the accuracy of its descriptions; all theories simplify and are therefore do not describe reality with great accuracy. Here it is important once again to distinguish between reasonable simplifications and assumptions that are simply false. Sometimes false assumptions can yield predictions that are consistent with one set of evidence (particularly when a researcher, consciously or not, designs the theory to fit some familiar evidence), but false assumptions are not likely to be consistent with new evidence. Predictions for new observations are more likely to be correct, and the goal of generalization is best served, when the assumptions of a theory are accurate descriptions, or at least reasonable simplifications. From this perspective, both assumptions (when they are observable) and predictions are fair game for testing.

The need for testing is exacerbated by the unfortunate emphasis that some rational-choice advocates place on their ability to generate non-obvious hypotheses (Geddes 2003, 193-198; Lave and March 1993, 67-73). It is easy to understand why they do this. On the one hand, a capacity to generate obvious or trivial hypotheses is of little value. On the other hand, the great breakthroughs in other sciences involve ideas that were not obvious before, such as the insights that the Earth is spherical, that all life evolved from primitive organisms, and that time slows down as one approaches the speed of light. But the fact that great scientific breakthroughs were not obvious does not logically imply that all non-obvious propositions are breakthroughs. (One would think that scholars who place so much emphasis on logic would have noticed this.) In fact, breakthroughs are breakthroughs not because they were not obvious, but because they fit a great deal of evidence in spite of initial expectations that they would not (Kuhn 1996, ; Lakatos 1970, ; Lakatos 1978). Common sense dictates that the less obvious an idea is, the more evidence is required to support it. As Carl Sagan was fond of saying, "Extraordinary claims require extraordinary evidence." Sometimes

an implausible idea improves after many revisions and eventually becomes an insight. But until that happens, counterintuitive propositions that do not fit the available evidence should be regarded as highly tentative and provisional.¹⁵

It is therefore especially tragic that the approach that tends to generate hypotheses most in need of testing has done so little testing and has done that little so poorly, as (Green and Shapiro 1994) have argued. However, as Green and Shapiro recognize, it takes time for a new body of formal theory to develop to a point where it is ready for testing. Their critique was directed at areas of rational-choice theorizing that have been active for several decades. Formal theories of democratization are relatively new, so it may be premature to expect much testing. Perhaps the fairest judgment is that no one should be surprised that the testing in this area has been inadequate.

Much of the theorizing in this approach uses such general concepts that it is not even feasible to test its consistency with evidence. Testing is problematic because no real-world cases can satisfy the highly restrictive assumptions of the model. Understood properly, these models predict something like "*If* there is one rich ruling actor and one poor actor, and *if* the ruler sets the tax rate and distributes assets, and *if* the poor would confiscate all the assets of the rich in a democracy, and *if* [. . .], *then* the greater inequality is, the less likely a democratic transition is." If we could find a case that satisfied all the "if" clauses, we could test the prediction fairly. But if there are three rich actors or fifteen poor actors or two middle classes or the tax rate is sticky or the poor would not confiscate many assets, etc., then the model makes no prediction and cannot be tested. The modeler could permit a test with imperfectly fitting cases on the grounds that the model makes only reasonable simplifications, so real-world cases match the model's premises well enough. But if the model were to fail the test, the modeler would be justified in crying foul on the grounds that a test using those cases was not fair after all. Whether a test is fair or not hinges on whether the assumptions of the

model were reasonable simplifications or oversimplifications. And this, in turn, hinges on whether the model captures the most important features or not, which is the empirical question that the test is supposed to answer (Morton 1999, 37-41). It thus becomes very hard to say whether a model failed a test because the test was inappropriate or because the model was oversimplified or wrong. By the same token, if a model passes an empirical test, it is very hard to know whether the test really confirms the model or the modeler simply got lucky (abetted by unrigorous testing).

Chapter 7 will lay out more comprehensive criteria for rigorous testing. However, the inadequacy of testing of formal theories is obvious in several ways that need not wait for the next chapter. First, some of the works in question are purely theoretical; they do not even attempt tests. The pure, untested, theories include those by Marks (1992), Sutter (1995), Acemoglu and Robinson (2001), and Zak and Feng (2003). These authors rarely even mention specific countries or events.

Second, most of the works that do refer to cases and evidence do so selectively, using the evidence to illustrate the theory rather than to test it. This is the case for all the works that are models rather than theories. The positional school has relied more on modeling specific cases as it has grown. Colomer modeled Spain (1991) and several post-communist cases (2000), Zielinski modeled Poland (1995) and Poland and Hungary (1999), Lohmann (1994) modeled East Germany, Weingast (1997) modeled the Glorious Revolution in England, Crescenzi (1999) modeled Hungary and Brazil, and Przeworski (1991) was obviously, if not always explicitly, inspired by transitions in Eastern Europe. These models have served to increase the plausibility of the approach, but have not been well integrated. The economic school relies more on general theories than models of cases, with the exception of Cohen (1994) on Brazil and Chile and Rosendorff (2001) on South Africa. In some respects one suspects that Feng and Zak's theory (1999) may have been inspired by East Asian experiences such as those of South Korea and Taiwan, due to the important role attributed to

student protests, and that Boix (2003) drew on the exceptional experiences of both the Middle East and his native Catalan region of Spain; but for the most part the economic school has been devoted to the development of general theory rather than applying it to specific cases.

A serious test would gather evidence for all the cases that satisfy the assumptions of the theory systematically, or at least a representative sample of them, thereby creating the risk that some of the evidence might turn out to be inconsistent with the predictions of the theory. Selection of only cases that are consistent with the theory, and which therefore are useful as illustrations, thwarts the goal of testing. Works in this category include Przeworski 1991 and Weingast 1997. Those who model specific events (listed in the previous paragraph) do essentially the same thing because there is virtually no chance that their model will not fit their case--or at least the details of the case that they choose to present. For these models, the examination of evidence is no more "testing" than the presentation of evidence in comparative histories; in fact, it is less rigorous because the evidence presented is far thinner, and therefore more selective.¹⁶

The third inadequacy of testing in this approach is that the small number of formal studies that do analyze systematic evidence tend to subject the evidence only to unchallenging tests. Often the scholar tests just one prediction of the model, even when the model yields other potentially testable implications, and bases the test on indicators that are only very indirect measures of the concepts in the model. Such tests neither confirm other predictions of the model nor disconfirm any competing hypotheses that are consistent with the same evidence. Swaminathan (1999), for example, makes predictions that depend on how risk-acceptant government and opposition are, how high the costs of conflict are, how much each actor discounts the future, and other factors. But his test merely confirms that the association between the incidence of democratic transitions and a rather crude indicator of "power parity" between government and opposition is not attributable to chance.

Crescenzi (1999) develops an elaborate model in which, after the regime liberalizes, the opposition must decide whether to cooperate, negotiate, or radicalize. The outcome could be a pacted transition, narrowed authoritarianism, or a revolutionary transition, depending on whether hardliners or softliners dominate the regime and the regime's ability to prevail in a conflict with the opposition. Lacking adequate data to model all of the possible paths, Crescenzi abandoned the details of his model's predictions and settled for testing the proposition that when the regime's recent acts of liberalization outnumber its acts of repression, the opposition is more likely to engage in peaceful protest (which he confirmed for Hungary 1948-1982 but could not confirm for Brazil 1964-1982). Feng and Zak (1999) claim that their theory is confirmed after they show that in 59 developing countries from 1970 to 1992, transitions to democracy are more likely in countries with higher income equality and either a high per capita GDP or a high level of education. Needless to say, these same results are compatible with many theories. In some respects, the most rigorous testing is the experimentation by Gould and Maggio (2003). However, the relevance of their experiments to regime change is debatable, since the subjects were U.S. college students who were asked to make decisions about monetary payoffs; the connection between these small payoffs and much larger economic stakes, and between economics and regime preferences, was simply assumed.

The most thorough testing of a formal theory has been done by Carles Boix in his *Democracy and Redistribution* (Boix 2003, chapter 2). His tests, which cover 12-20 countries from 1950 to 1990 and 208 states that existed at any time from 1850 to 1980, equivalent to approximately 6,500 country-years, are as general as available data permit. They include tests of not one hypothesis from his theory, but several: that equality favors democracy, that oil producers tend to be less democratic, and that it was industrialization rather than per capita GDP that drove the first wave of democratization. He controls for several competing explanations in addition to vindicating the

predictions of his own theory, and tests thick implications of his theory with process-tracing over centuries of Swiss and U.S. history. Chapters 7 and 10 will examine the rigor of Boix's tests in more detail, but for now it suffices to say that *Democracy and Redistribution* is a resounding exception to the rule that formal theories have not been tested.

It is natural to hope that as formal theories come to be tested more frequently and thoroughly, their kinks will be progressively straightened out until they jibe with common sense and generate predictions that are not only unexpected, but also consistent with most of the available evidence. However, there is nothing about repeated encounters between this theory (or any theory) and evidence that guarantees such an outcome. The sobering reality is that inconsistent evidence offers no guidance about where the theory went wrong or which proposition should be modified to make it right (the "Duhem-Quine Hypothesis").¹⁷ However, because scholars have faith in the potential of their approach, they tend to revise only the peripheral assumptions of the theory, not to question its core, a practice analogous to the "last hired, first fired" norm.

Formal theory's emphasis on integration therefore works against empirical generality in several ways. While the core theory is being developed, it is insulated from potentially disconfirming evidence, which is considered irrelevant until the theory becomes complex enough to approximate reality. Modeling begins to bridge the gaps between theory and reality, but with the goal of vindicating core assumptions rather than questioning them. And when the theory eventually confronts systematic disconfirming evidence, the core theory is still safeguarded; only peripheral assumptions come under fire. It is only when a great deal of evidence inconsistent with a theory accumulates and a more promising alternative approach emerges that scholars begin to question the core assumptions rather than the peripheral propositions, and the theory is replaced in a "scientific revolution" or "paradigm shift" (Kuhn 1996, ; Lakatos 1970). Even then, according to Kuhn, few scholars change

their minds. A paradigm dies when its adherents retire and are replaced by the followers of a different paradigm. Integration is an essential quality of good theory, but it forces a steep tradeoff with generality.

Conclusion

The value of the theorizing in formal models of democratization has not been integration so much as posing new questions. By using strict logic or mathematics to derive implications, they reveal the fallacies and non sequiturs in non-formal theories. Formal theory has repeatedly identified problematic assumptions that cannot be taken for granted. Several of these issues are relevant for democratization. For example, several approaches treat democratization as though it were a non-strategic process. This is especially true of approaches without actors, such as those that emphasize socioeconomic modernization, mass political culture, economic crisis, divided societies, or some varieties of international diffusion. Even when these approaches refer to actors, their choices are treated as reflexive responses: involuntary reactions to objective conditions. Rational-choice models focus attention on the intentions of actors who have choices. What these actors choose, and why, becomes a central part of the explanation for the outcome. In a sense, their choices are completely determined by the situation in which they find themselves; but because the “correct” choices are not obvious, a theory is needed to explain why actors choose as they do. Non-formal models tend to overlook such issues.

Formal models and formal theories of democratization constitute a distinctive approach because they do a relatively good job of satisfying one criterion for good theory. They come closer to achieving some degree of theoretical integration than checklists, case studies, comparative histories, or (as we shall see later) statistical analyses. The degree of integration may be less than one would hope; in fact, there are two completely distinct schools of theory within this approach, one of

which (the positional school) is not integrated well internally. Nevertheless, there is more integration here than elsewhere. But like the other approaches--small-sample qualitative analyses and large-sample quantitative analyses--formal theory does a relatively poor job of satisfying two criteria for good theory. This approach is perhaps the thinnest of all, and most of it (Boix 2003 excepted) cannot claim to approach generality.

Table 6.1: Formal Theories of Democratization

I. The positional school

A. Theories of liberalization

(Crescenzi 1999, ; Lohmann 1994, ; Marks 1992, ; Przeworski 1986, ; Przeworski 1991, ; Zielinski 1995)

B. Theories of transition and survival

1. Unitary rulers vs. unitary opposition

(Sutter 1995, ; Swaminathan 1999)

a. With a foreign actor: (Sutter 1995)

2. With one actor subdivided

a. Unitary opposition vs. divided rulers

1) Hardliners, Softliners, Opposition

(Crescenzi 1999, ; Zielinski 1999)

b. Unitary rulers vs. divided opposition

1) Defender, Challenger, Mass Public: (Casper and Taylor 1996)

2) Sovereign, Citizens A, Citizens B: (Weingast 1997)

3. With both actors subdivided

a. Hardliners, reformers, radicals, moderates: (Przeworski 1991)

b. Radical and moderate factions of hardliners, softliners, and opposition:
(Colomer 1991, ; Colomer 2000)

II. The economic school

A. About breakdown, involving radical and moderate factions of Left and Right:

(Cohen 1994)

B. All about transitions and/or survival, all involving two actors, rich and poor:

(Acemoglu and Robinson 2001, ; Boix 2003, ; Feng and Zak 1999, ; Gould and Maggio 2003, ; Rosendorff 2001, ; Zak and Feng 2003)

1. With a third actor (a middle class) and consideration of capital mobility: (Boix 2003)

Notes

I am grateful to Annabella España Nájera for research assistance with this chapter, funded by the Kellogg Institute.

1. There are many other possible ways to categorize works in this approach—by the kind of game being played, the number of actors, the options available to them, the stage of democratization, the nature of rationality, the degree of formalization, and so on. I have chosen the nature of the actors and their goals as the most fundamental divide because it is the most substantively interesting difference and it groups together the best-integrated set of works—those assuming that economic interests are at stake.
2. In a somewhat different analysis, Zielinski (1999) applied reasoning like Przeworski's to different actors: the Soviet Union and a Government, instead of hardliners and softliners. This was probably a useful shift in focus for understanding his case—Poland in the 1980s. His conclusion was that the less likely a Soviet intervention was, the more likely an opening became.
3. Although the logic is the same, the authors use different terminologies. Sutter (1995) calls the ruler “the dictator,” defines his choices as “fighting” or “abdicating,” and allows the opposition to “rebel” or “accept” dictatorship. Swaminathan (1999) calls the ruler “the government,” defines its choices as “negotiating” or not, and allows the opposition to “mobilize” or not.
4. If I were to push the genealogy analogy, these models would be cousins of a nephew of the Sutter and Swaminathan models.
5. Zielinski does not allow for the opposition to prevail in a violent struggle, so this model rules out revolutionary transitions.
6. Casper and Taylor's model is only very loosely formalized. They present diagrams that resemble extended-form games, but they never present an actual game with payoffs or preference orderings or sophisticated strategies. Some of their fundamental assumptions remain implicit.
7. Weingast claims more general applicability of his model, but since he claims that it reveals the critical nature of pacts in transitions and defines democratic consolidation as the achievement of the kind of coordination he describes, it is appropriate to discuss it in this section.
8. Przeworski later describes a consolidation game in which the incumbent must decide whether or not to accept an election loss. He argues that this choice depends on constitutional rules that affect the incumbent's chances of returning to power.
9. Boix (2003, 9) makes the same point, but I arrived at this conclusion independently.
10. The theory predicts that overspending on education or police would hurt growth and therefore slow democratization, but the authors claim to have shown elsewhere that governments generally underspend on both. In other words, typically spending more on police would spur growth and therefore accelerate democratization.

11. The original inspirations for these models, however, appear to be economic models of redistribution such as {Alesina, 1994 #243} or {Persson, 1994 #242}, which were not intended to explain regime change.
12. Rosendorff also assumed that the elite has the option of investing its capital abroad rather than in the domestic economy. The elite can use this as leverage to keep the tax rate below 100 percent, even in a democracy. He also says that the elite can spend on "maintaining control" rather than on education and police, specifically. Because these modifications are mostly irrelevant to the predictions about regime transition, I have omitted them from the main discussion.
13. Boix (2003) never cites Zak and Feng (2003), or even Feng and Zak (1999), so he was either unaware of their work or did not wish to share credit with them for a couple of innovations.
14. In an appendix, Boix (63-64) develops two alternative reasons for the association between development and democracy. It may be that a democratic government would keep taxes low to prevent taxation from dampening productivity, or it may be that taxation is subjectively less onerous to the wealthy at higher levels of income. However, he places more emphasis on the role of asset specificity.
15. The exception to this is that counterintuitive propositions that are rigorously derived from propositions that are known to be true should be seriously entertained until evidence can be gathered to test them. This exception applies to mathematics and some natural sciences that have developed many laws that are known to be true. It does not apply to formal theories in political science, which are derived from propositions that are usually oversimplified or false.
16. Casper and Taylor (1996) belong in this category as well despite their large sample (24 countries) because ten of these cases had to be assigned to some "alternate" path that differed from the original theory. They adjusted the model to fit the cases.
17. Diermeir (1996 @63) states the Duhem-Quine Hypothesis as "(1) A theory is accepted or rejected as a whole, not through the acceptance or rejection of particular components of the theory. (2) A rejection of a theory based on a critical experiment is impossible."

References

- Acemoglu, Daron, and James A. Robinson. 2001. A Theory of Political Transitions. *The American Economic Review* 91 (4):938-63.
- Alesina, Alberto, and Dani Rodrik. 1994. Distributive Politics and Economic Growth. *Quarterly Journal of Economics* 112:465-490.
- Boix, Carles. 2003. *Democracy and Redistribution*. New York: Cambridge University Press.
- Casper, Gretchen, and Michelle M. Taylor. 1996. *Negotiating Democracy: Transitions from Authoritarian Rule*. Pittsburg: University of Pittsburg Press.
- Cohen, Youssef. 1994. *Radicals, Reformers, and Reactionaries: The Prisoner's Dilemma and the Collapse of Democracy in Latin America*. Chicago: University of Chicago Press.
- Colomer, M. Josep. 1991. Transitions By Agreement: Modeling the Spanish Way. *American Political Science Review* 85 (4):1283-1302.
- Colomer, M. Josep. 2000. *Strategic Transitions: Game Theory and Democratization*. Baltimore: The Johns Hopkins University Press.
- Crescenzi, Mark J.C. 1999. Violence and Uncertainty in Transitions. *Journal of Conflict Resolution* 43 (2):192-212.
- Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
- Feng, Yi, and Paul J. Zak. 1999. The Determinants of Democratic Transitions. *Journal of Conflict Resolution* 43 (2):162-77.
- Fudenberg, Drew, and Eric Maskin. 1986. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica* 54 (3):533-554.
- Geddes, Barbara. 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor, MI: University of Michigan Press.
- Gould, Andrew C., and Andrew K. Maggio. 2003. Political Regimes and Economic Development: A Framing-Effects Model and Experimental Data. Notre Dame, IN.
- Green, Donald P., and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven: Yale University Press.
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47 (2):263-291.
- Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago: University of Chicago Press.
- Lakatos, Imre. 1970. Falsification and the Methodology of Scientific Research Programmes. In *Criticism and the Growth of Knowledge*, edited by I. Lakatos, and A. Musgrave. Cambridge: Cambridge University Press.
- Lakatos, Imre. 1978. *The Methodology of Scientific Research Programmes: Philosophical Papers*. Edited by J. Worrall, and Gregory Currie. Vol. 1. Cambridge: Cambridge University Press.
- Lave, Charles A., and James G. March. 1993. *An Introduction to Models in the Social Sciences*. Lanham, MD: University Press of America.
- Lohmann, Susanne. 1994. The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91. *World Politics* 47 (1):42-101.
- Marks, Gary. 1992. Rational Sources of Chaos in Democratic Transitions. In *Reexamining Democracy: Essays in Honor of Seymour Martin Lipset*, edited by G. a. L. D. Marks. Newbury Park, CA: Sage Publications.
- Morton, Rebecca B. 1999. *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science*. Cambridge: Cambridge University Press.
- O'Donnell, Guillermo, and Philippe C. Schmitter. 1986. *Transitions from Authoritarian Rule: Tentative*

- Conclusions about Uncertain Democracies*. 5 vols. Vol. 5. Baltimore: The Johns Hopkins University Press.
- Persson, Torsten, and Guido Tabellini. 1994. Is Inequality Harmful for Growth? Theory and Evidence. *American Economic Review* 84 (3):600-621.
- Przeworski, Adam. 1986. Problems in the Study of Transition to Democracy. In *Transitions from Authoritarian Rule: Comparative Perspectives*, edited by G. O'Donnell, Philippe C. Schmitter, and Laurence Whitehead. Baltimore: The Johns Hopkins University Press.
- Przeworski, Adam. 1991. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Cambridge: Cambridge University Press.
- Rosendorff, B. Peter. 2001. Choosing Democracy. *Economics and Politics* 13 (1):1-29.
- Sutter, Daniel. 1995. Settling Old Scores: Potholes along the Transition from Authoritarian Rule. *Journal of Conflict Resolution* 39 (1):110-128.
- Swaminathan, Siddharth. 1999. Time, Power, and Democratic Transitions. *Journal of Conflict Resolution* 43 (2):178-91.
- Weingast, Barry R. 1997. The Political Foundations of Democracy and the Rule of Law. *American Political Science Review* 91 (2):245-63.
- Zak, Paul J., and Yi Feng. 2003. A Dynamic Theory of the Transition to Democracy. *Journal of Economic Behavior and Organization* 52:1-25.
- Zielinski, Jakub. 1995. The Polish Transition to Democracy: A Game-Theoretic Approach. *Archives Européennes de Sociologie* 36 (1):135-158.
- Zielinski, Jakub. 1999. Transitions from Authoritarian Rule and the Problem of Violence. *Journal of Conflict Resolution* 43 (2):213-228.