Chapter One

Statement of the Problem

1.1 WHY DO WE NEED A MATHEMATICAL THEORY IN HISTORY?

Why do some polities—chiefdoms and states of various kinds—embark on a successful program of territorial expansion and become empires? Why do empires sooner or later collapse? Historians and sociologists offer a great variety of answers to these and related questions. These answers range from very specific explanations focusing on unique characteristics of one particular polity to quite general theories of social dynamics. There has always been much interest in understanding history, but recently the theoretical activity in this area has intensified (Rozov 1997). Historical sociology is attempting to become a theoretical, mature science.

But why do historical sociologists use such a limited set of theoretical tools? *Theory* in social sciences usually means careful thinking about concepts and definitions. It is verbal, conceptual, and discursive. The theoretical propositions that are derived are *qualitative* in nature. Nobody denies the immense value of such theoretical activity, but it is not enough. There are also formal, mathematical approaches to building theory that have been applied with such spectacular success in physics and biology. Yet formalized theory employing mathematical models is rarely encountered in historical sociology (we will be reviewing some of the exceptions in later chapters).

The history of science is emphatic: a discipline usually matures only after it has developed mathematical theory. The requirement for mathematical theory is particularly important if the discipline deals with dynamic quantities (see the next section). Everybody is familiar with the paradigmatic example of classical mechanics. But two more recent examples from biology are the synthetic theory of evolution that emerged during the second quarter of the twentieth century (Ruse 1999), and the ongoing synthesis in population ecology (for example, Turchin 2003). In all these cases, the impetus for synthesis was provided by the development of mathematical theory.

Can something similar be done in historical sociology? Several attempts have been made in the past (e.g., Bagehot 1895; Rashevsky 1968), but they clearly failed to make an impact on how history is studied today. I think there are two major reasons explaining this failure. First, these attempts were inspired directly by successes in physical sciences. Yet physicists traditionally choose to deal with systems and phenomena that are very different from those in history. Physicists

tend to choose very simple systems with few interacting components (such as the solar system, the hydrogen atom, etc.) or systems consisting of a huge number of identical components (as in thermodynamics). As a result, very precise quantitative predictions can be made and empirically tested. But even in physical applications such systems are rare, and in social sciences only very trivial questions can be reduced to such simplicity. Real societies always consist of many qualitatively and quantitatively different agents interacting in very complex ways. Furthermore, societies are not closed systems: they are strongly affected by exogenous forces, such as other human societies, and by the physical world. Thus, it is not surprising that traditional physical approaches honed on simple systems should fail in historical applications.

The second reason is that the quantitative approaches typically employed by physicists require huge amounts of precisely measured data. For example, a physicist studying nonlinear laser dynamics would without further ado construct a highly controlled lab apparatus and proceed to collect hundreds of thousands of extremely accurate measurements. These data would then be analyzed using sophisticated methods on a high-powered computer. Nothing could be further from the reality encountered by a historical sociologist, who typically lacks data about many aspects of the historical system under study, while possessing fragmentary and approximate information about others. For example, one of the most important aspects of any society is just how many members it has. But even this kind of information usually must be reconstructed by historians on the basis of much guesswork.

If these two problems are the real reason why previous attempts failed, then some recent developments in natural sciences provide a basis for hope. First, during the last 20–30 years, physicists and biologists have mounted a concerted attack on complex systems. A number of approaches can be cited here: nonlinear dynamics, synergetics, complexity, and so on. The use of powerful computers has been a key element in making these approaches work. Second, biologists, and ecologists in particular, have learned how to deal with short and noisy data sets. Again, plentiful computing power was a key enabler, allowing such computer-intensive approaches as nonlinear model fitting, bootstrapping, and cross-validation.

There is another hopeful development, this time in social sciences. I am referring to the rise of quantitative approaches in history, or *cliometrics* (Williamson 1991). Currently, there are many investigators who collect quantitative data on various aspects of historical processes, and large amounts of data are already available in electronic form.

These observations suggest that another attempt at building and testing quantitative theories in historical sociology may be timely. If we achieve even partial success, the potential payoff is so high that it warrants making the attempt. And there are several recent developments in which application of modeling and quantitative approaches to history have already yielded interesting insights.

1.2 HISTORICAL DYNAMICS AS A RESEARCH PROGRAM

Many historical processes are *dynamic*. Generally speaking, *dynamics* is the scientific study of any entities that change with time. One aspect of dynamics deals with a phenomenological description of temporal behaviors—trajectories (this is sometimes known as kinematics). But the heart of dynamics is the study of mechanisms that bring about temporal change and explain the observed trajectories. A very common approach, which has proved its worth in innumerable applications, consists of taking a holistic phenomenon and mentally splitting it up into separate parts that are assumed to interact with each other. This is the dynamical systems approach, because the whole phenomenon is represented as a *system* consisting of several interacting *elements* (or *subsystems*, since each element can also be represented as a lower-level system).

As an example, consider the issue raised at the very beginning of the book. An empire is a dynamic entity because various aspects of it (the most obvious ones being the extent of the controlled territory and the number of subjects) change with time: empires grow and decline. Various explanations for imperial dynamics address different aspects of empires. For example, we may be concerned with the interacting processes of surplus product extraction and warfare (e.g., Tilly 1990). Then we might represent an empire as a system consisting of such subsystems as the peasants, the ruling elite, the army, and perhaps the merchants. Additionally, the empire controls a certain territory and has certain neighboring polities (that is, there is a higher-level system—or *metasystem*—that includes the empire we study as a subsystem). In the dynamical system's approach, we must describe mathematically how different subsystems interact with each other (and, perhaps, how other systems in the metasystem affect our system). This mathematical description is the model of the system, and we can use a variety of methods to study the dynamics predicted by the model, as well as attempt to test the model by comparing its predictions with the observed dynamics.

The conceptual representation of any holistic phenomenon as interacting subsystems is always to some degree artifical. This artificiality, by itself, cannot be an argument against any particular model of the system. All models simplify the reality. The value of any model should be judged only against alternatives, taking into account how well each model predicts data, how parsimonious the model is, and how much violence its assumptions do to reality. It is important to remember that there are many examples of very useful models in natural sciences whose assumptions are known to be wrong. In fact, all models are by definition wrong, and this should not be held against them.

Mathematical models are particularly important in the study of dynamics, because dynamic phenomena are typically characterized by nonlinear feedbacks, often acting with various time lags. Informal verbal models are adequate for generating predictions in cases where assumed mechanisms act in a linear and additive fashion (as in trend extrapolation), but they can be very misleading when we deal with a system characterized by nonlinearities and lags. In general, nonlinear dynamical systems have a much wider spectrum of behaviors than could be imagined by informal reasoning (for example, see Hanneman et al. 1995). Thus,

a formal mathematical apparatus is indispensable when we wish to rigorously connect the set of assumptions about the system to predictions about its dynamic behavior.

1.2.1 Delimiting the Set of Questions

History offers many puzzles and somehow we must select which of the questions we are going to address in this research program. I chose to focus on territorial dynamics of polities, for the following reasons. Much of recorded history is concerned with territorial expansion of one polity at the expense of others, typically accomplished by war. Why some polities expand and others fail to do so is a big, important question in history, judging, for example, by the number of books written about the rise and fall of empires. Furthermore, the spatiotemporal record of territorial state dynamics is perhaps one of the best quantitative data sets available to the researcher. For example, the computer-based atlas CENTENNIA (Reed 1996) provides a continuous record of territorial changes during 1000–2000 C.E. in Europe, Middle East, and Northern Africa. Having such data is invaluable to the research program described in this book, because it can provide a *primary data set* with which predictions of various models can be compared.

The *dynamic* aspect of state territories is also an important factor. As I argued in the previous section, dynamic phenomena are particularly difficult to study without a formal mathematical apparatus. Thus, if we wish to develop a mathematical theory for history, we should choose those phenomena where mathematical models have the greatest potential for nontrivial insights.

Territorial dynamics is not the whole of history, but it is one of the central aspects of it, in two senses. First, we need to invoke a variety of social mechanisms to explain territorial dynamics, including military, political, economic, and ideological processes. Thus, by focusing on territorial change we are by no means going to be exclusively concerned with military and political history. Second, characteristics of the state, such as its internal stability and wealth of ruling elites, are themselves important variables explaining many other aspects of history, for example, the development of arts, philosophy, and science.

1.2.2 A Focus on Agrarian Polities

There are many kinds of polities, ranging from bands of hunter-gatherers to the modern postindustrial states. A focus on particular socioeconomic formation is necessary if we are to make progress. The disadvantages of industrial and postindustrial polities are that the pace of change has become quite rapid and the societies have become very complex (measured, for example, by the number of different professions). Additionally, we are too close to these societies, making it harder for us to study them objectively. The main disadvantage of studying hunter-gatherer societies, on the other hand, is that we have to rely primarily on archaeological data. Agrarian societies appear to suffer the least from these two disadvantages: throughout most of their history they changed at a reasonably slow pace, and we have good historical records for many of them. In fact, more

than 95% of recorded history is the history of agrarian societies. As an additional narrowing of the focus for this book, I will say little about nomadic pastoralist societies and leave out of consideration thalassocratic city-states (however, both kinds of polities are very important, and will be dealt with elsewhere).

This leaves us still with a huge portion of human history, roughly extending from –4000 to 1800 or 1900 C.E.,[1] depending on the region. One region to which I will pay much attention is Europe during the period 500–1900 C.E., with occasional excursions to China. But the theory is meant to apply to all agrarian polities, and the aim is to test it eventually in other regions of the world.

1.2.3 The Hierarchical Modeling Approach

There is a heuristic "rule of thumb" in modeling dynamical systems: do not attempt to encompass in your model more than two hierarchical levels. A model that violates this rule is the one that attempts to model the dynamics of both interacting subsystems within the system *and* interactions of subsubsystems within each subsystem. Using an individual-based simulation to model interstate dynamics also violates this rule (unless, perhaps, we model simple chiefdoms). From the practical point of view, even powerful computers take a long time to simulate systems with millions of agents. More importantly, from the conceptual point of view it is very difficult to interpret the results of such a multilevel simulation. Practice shows that questions involving multilevel systems should be approached by separating the issues relevant to each level, or rather pair of levels (the lower level provides mechanisms, one level up is where we observe patterns).

Accordingly, in the research program described in this book I consider three classes of models. In the first class, individuals (or, perhaps, individual households) interact together to determine group dynamics. The goal of these models is to understand how patterns at the group level arise from individual based mechanisms. In the second class, we build on group-level mechanisms to understand the patterns arising at the polity level. Finally, the third class of models addresses how polities interact at the interstate level. The greatest emphasis will be on the second class of models (groups–polity). I realize that this sounds rather abstract at this point; in particular, what do I mean by "groups"? The discussion of this important issue is deferred until chapter 3. Also, I do not wish to be too dogmatic about following the rule of two levels. When we find it too restrictive, we should break it; the main point is not to do it unless really necessary.

1.2.4 Mathematical Framework

The hard part of theory building is choosing the mechanisms that will be modeled, making assumptions about how different subsystems interact, choosing functional forms, and estimating parameters. Once all that work is done, obtaining model predictions is conceptually straightforward, although technical, laborious,

---

[1]Negative sign refers to years B.C.E.

and time consuming. For simpler models, we may have analytical solutions available (to solve a model analytically means to derive a formula that gives a precise solution for all parameter values). However, once the model reaches even a medium level of complexity we typically must use a second method: solving it numerically on the computer. A third approach is to use agent-based simulations (Kohler and Gumerman 2000). These ways of obtaining model predictions should not be considered as strict alternatives. On the contrary, a mature theory employs all three approaches synergistically.

Agent-based simulation (ABS), for example, is a very powerful tool for investigating emerging properties of a society consisting of individuals who are assumed to behave in a certain way (by redefining agents to mean groups of individuals or whole polities, we can also use this approach to address higher-level issues). Agent-based models are easily expandable, we can add various stochastic factors, and in general model any conceivable mechanisms. In principle, it is possible to build a theory by using only agent-based simulations. In practice, however, a sole emphasis on these kinds of models is a poor approach. One practical limitation is that currently available computing power, while impressive, is not infinite, putting a limit on how much complexity we can handle in an agent-based simulation. More importantly, ABSs have conceptual drawbacks. Currently, there is no unified language for describing ABSs, making each particular model opaque to everybody except those who are steeped in the particular computer language the model is implemented in. Small details of implementation may result in big differences in the predicted dynamics, and only in very rare cases do practitioners working with different languages bother to cross-translate their ABS (for a rare exception, see Axelrod 1997). And, finally, the power of ABSs is at the same time their curse: it is too easy to keep adding components to these models, and very soon they become too complex to understand.

The more traditional language for modeling dynamical systems, based on differential (or difference) equations, has several advantages. First, it has been greatly standardized, so that a model written as a system of differential equations is much easier to grasp than the computer code describing the same assumptions. This, of course, assumes that the person viewing the model has had much experience with such equations, which unfortunately is not the case with most social scientists, or even biologists, for that matter. Still, one may hope that the level of numeracy in nonphysical sciences will increase with time, and perhaps this book will be of some help here. Second, analytical results are available for most simple or medium-complexity models. Even if we do not have an explicit analytical solution (which is the case for most nonlinear models), we can obtain analytical insights about qualitative aspects of long-term dynamics predicted by these models. Third, numerical methods for solving differential models have been highly standardized. Thus, other researchers can rather easily check on the numerical results of the authors. To sum up, differential (difference) equations provide an extremely useful common language for theory building in dynamical applications.

Note that I am not arguing against the use of ABSs. In fact, I find the recently proposed agenda for doing social science from the bottom up by growing

artifical societies (Epstein and Axtell 1996) extremely exciting (for an excellent volume illustrating the strength of this approach when applied to real problems in the social sciences, see Kohler and Gumerman 2000). Rather, I suggest that the ABS should always be supplemented by other approaches, which may lack the power of ABSs, but are better at extracting, and communicating, the important insights from the chaos of reality. The best approach to building theory is the one that utilizes all the available tools: from pencil-and-paper analysis of models to numerical solutions to agent-based simulations.

1.3 SUMMARY

To summarize the discussion in this introductory chapter, here is my proposal for a research program for theory building in historical dynamics.

- Define the problem to be addressed: the territorial dynamics of agrarian polities. The main questions are, why do some polities at certain times expand? And why do they, at other times, contract, or even completely disappear? More luridly, what are the causal mechanisms underlying the rise and demise of empires?
- Identify the primary data set: the spatiotemporal record of territorial dynamics within a certain part of the world and a certain period of time. The data set serves as the testing bed for various mechanistic theories. The success of each theory is measured by how well its predictions match quantitative patterns in the primary data.
- Identify a set of hypotheses, each proposing a specific mechanism, or a combination of mechanisms, to explain territorial expansion/contraction of polities. Many of these hypotheses have already been proposed, others may need to be constructed de novo. The list of hypotheses does not have to be exhaustive, but it should include several that appear most likely, given the present state of knowledge. Hypotheses also do not need to be mutually exclusive.
- Translate all hypotheses in the list into mathematical models. Typically, a single hypothesis will be translated into a spectrum of models, using alternative assumptions about functional forms and parameter values.
- Identify secondary data. These are the data that we need for each specific hypothesis and its associated spectrum of models. For example, if a hypothesis postulates a connection between population growth and state collapse, then we need data on population dynamics. Secondary data provide the basis for auxiliary tests of hypotheses (in addition to tests based on the primary data). Thus, predictions from a hypothesis based on population dynamics should match the observed patterns in the population data. On the other hand, a hypothesis based on legitimacy dynamics does not need to predict population data also; instead, its predictions should match temporal fluctuations of legitimacy.

- Solve the models using appropriate technology (that is, analytical, numerical, and simulation methods). Select those features of the models' output where there is a disagreement among hypotheses/models, and use the primary data set to determine which hypothesis predicts this aspect better than others. Take into account the ability of each hypothesis to predict the appropriate secondary data, how parsimonious is the model into which the hypothesis is translated, and any degree of circularity involved (for example, when the same data are used for both parameter estimation and model testing). Make a tentative selection in favor of the model (or models) that predicts various features of the data best with the least number of free parameters.
- Repeat the process, by involving other hypotheses and by locating more data that can be used to test various models.

Clearly this is a highly idealized course of action, which sounds almost naive in its positivistic outlook. In practice, it is unlikely that it will work just as described above. Nevertheless, there is a value in setting the goal high. The rest of the book presents a deliberate attempt to follow this research program. As we shall see, reality will intrude in a number of sobering ways. Yet I also think that the results, while failing to achieve the lofty goals set out above, prove to be instructive. But this is for the readers to judge.