HISTORICAL DYNAMICS

Why States Rise and Fall

PETER TURCHIN

Copyrighted material

Historical Dynamics

Why States Rise and Fall

Peter Turchin

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Copyrighted material

Copyright © 2003 by Princeton University Press
Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 6 Oxford Street,
Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

All Rights Reserved

First published in 2003

First paperback printing, 2018

Paper ISBN 978-0-691-18077-9
Cloth ISBN: 978-0-691-11669-3

Library of Congress Control Number 2003110656

British Library Cataloging-in-Publication Data is available

The publisher would like to acknowledge Peter Turchin for providing
the camera-ready copy from which this book was printed

Printed on acid-free paper. ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Copyrighted material

Contents

Copyrighted material

Copyrighted material

Copyrighted material

List of Figures

Copyrighted material

Copyrighted material

List of Tables

Copyrighted material

Preface

Many historical processes are dynamic: growth and decline of populations, territorial expansion and contraction of empires, trends in political centralization/decentralization, and the spread of world religions, to name just a few examples. A general approach to studying dynamical systems is to advance rival hypotheses based on specific mechanisms, translate the hypotheses into mathematical models, and contrast model predictions with empirical patterns. Mathematical modeling is a key ingredient in this research program because quantitative dynamical phenomena, often affected by complex feedbacks, cannot be fully understood at a purely verbal level. Another important ingredient is the full use of statistical techniques (such as time-series analysis) for quantitative and rigorous comparison between model-predicted and observed patterns. This general approach has proved to be extremely successful in natural sciences. Can it be instrumental in increasing our understanding of historical processes?

Historical Dynamics is an attempt to answer this question. The specific problem chosen for analysis is the territorial dynamics of agrarian states. In other words, can we understand why some polities at certain times expand, while at other times they contract? The advantage of focusing on territorial expansion/contraction is that we have reasonably accurate empirical data on this aspect of historical dynamics (historical atlases). The focus on agrarian polities is motivated by the extent of empirical material (roughly, from the third millennium B.C.E. to 1800 C.E.) and the greater simplicity of these societies compared to modern ones, potentially making them easier to understand and model.

Although the main focus of the book is on territorial dynamics, it is clear that the ability (or inability) of states to expand depends very much on their internal characteristics. Thus, in order to understand how and why states expand and contract, we need to study military, economic, demographic, ethnological, and ideological aspects of social dynamics. I consider four sociological theories potentially explaining territorial dynamics. The first is the geopolitical model of Randall Collins. This theory has been very clearly formulated and requires minimal work to translate into a mathematical model. The second one, by contrast, is an original development. Starting from ideas of the fourteenth century Arabic thinker Ibn Khaldun and recent developments in sociobiology, I advance a theory attempting to explain why the capacity for collective action may vary among different societies. The third theory addresses the issue of ethnic assimilation/religious conversion dynamics. Finally, the fourth theory focuses on the interaction between population dynamics and sociopolitical stability. The connection between population growth and state breakdown is based on the demographic-structural model of Jack Goldstone (another well-formulated theory that is easily

Copyrighted material

translated into a dynamical model). To this model, I add the feedback mechanism, postulating how state breakdown and resulting sociopolitical instability negatively affect population numbers. The four theories address somewhat different aspects of historical dynamics, and thus logically are not mutually exclusive. However, alternative hypotheses about particular empirical patterns can be derived from them and tested with data. I present several such empirical tests.

ACKNOWLEDGMENTS

Many people provided extensive comments on previous book drafts or draft chapters. I am particularly indebted to Tom Hall, Jack Goldstone, Sergey Nefedov, and the anonymous reviewer who read the whole draft and provided numerous and truly excellent comments and criticisms. I also wish to express my deep gratitude to Marc Artzrouni, Robert Boyd, Christopher Chase-Dunn, Randall Collins, Lev Ginzburg, Robert Hanneman, John Komlos, and Nikolai Rozov for their comments on various parts of previous drafts. Many thanks to Svetlana Borinskaya, Andrey Korotayev, and other members of the Social Evolution group in Moscow for constructive critique and general encouragement. I am grateful to Marc Artzrouni for providing the graphical output of his simulation model for Figure 2.4. Finally, I wish to thank Jennifer Slater for excellent copyediting, Kathy Tebo for help with typing and proofreading, and Mirko Janc for his "TeXpert" typesetting.

Copyrighted material

Historical Dynamics

Why States Rise and Fall

Copyrighted material

Copyrighted material

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

Copyrighted material

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.

Copyrighted material

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,

Copyrighted material

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

Copyrighted material

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.

Copyrighted material

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

Copyrighted material

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.

Copyrighted material

- 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.

Copyrighted material

Chapter Two

Geopolitics

Geopolitics, in the narrow sense that I use in this book, concerns itself with the spatial aspects of historical dynamics. There are two major kinds of mechanisms invoked in geopolitical models: the ability to project power at distance, and the effect of spatial position. Thus, geopolitics is a natural place to start in my review of theories for territorial dynamics of states. Additionally, it is one of the best-theorized areas in historical sociology (for example, Collins 1978, 1986, 1995), and perhaps enjoys the greatest number of already existing formal models (see below). However, my argument in this chapter indicates that geopolitical models (in the narrow sense) are insufficient for the explanation of empirical patterns; in particular, they fail to account for a sustained decline of formerly powerful and territorially extensive polities.

2.1 A PRIMER OF DYNAMICS

Rather than starting immediately with geopolitical models, I first review some basic facts about general kinds of behaviors that can be exhibited by dynamical systems. Although these facts are fairly elementary, they are worth discussing, because (1) they may not be well known to people lacking extensive experience with dynamic models, and (2) it gives me a chance to introduce a simple classification scheme, to which I can then refer throughout the book. An excellent introduction to dynamical social systems can be found in Fararo (1989).

2.1.1 Boundless Growth

The simplest possible dynamics is linear growth, obeying the differential equation

$$\dot{X} = c \tag{2.1}$$

Here X is the variable that is changing, $\dot{X}$ is the rate of change of X (often written as dX/dt), and c is a constant. A familiar application of this model is Newton's first law, in which X is the position of a body and c a constant velocity. (The first law states that, in the absence of any forces acting on the body, it will move with constant velocity.) The solution of this equation is $X(t) = X_0 + ct$, where t is time and X_0 is the initial position of the body, $X(0) = X_0$. The solution says that X will change linearly with time (Figure 2.1a), and that the rate of change is c. Thus, I refer to this kind of dynamics as *linear growth* (or decline, if c is negative).

Copyrighted material

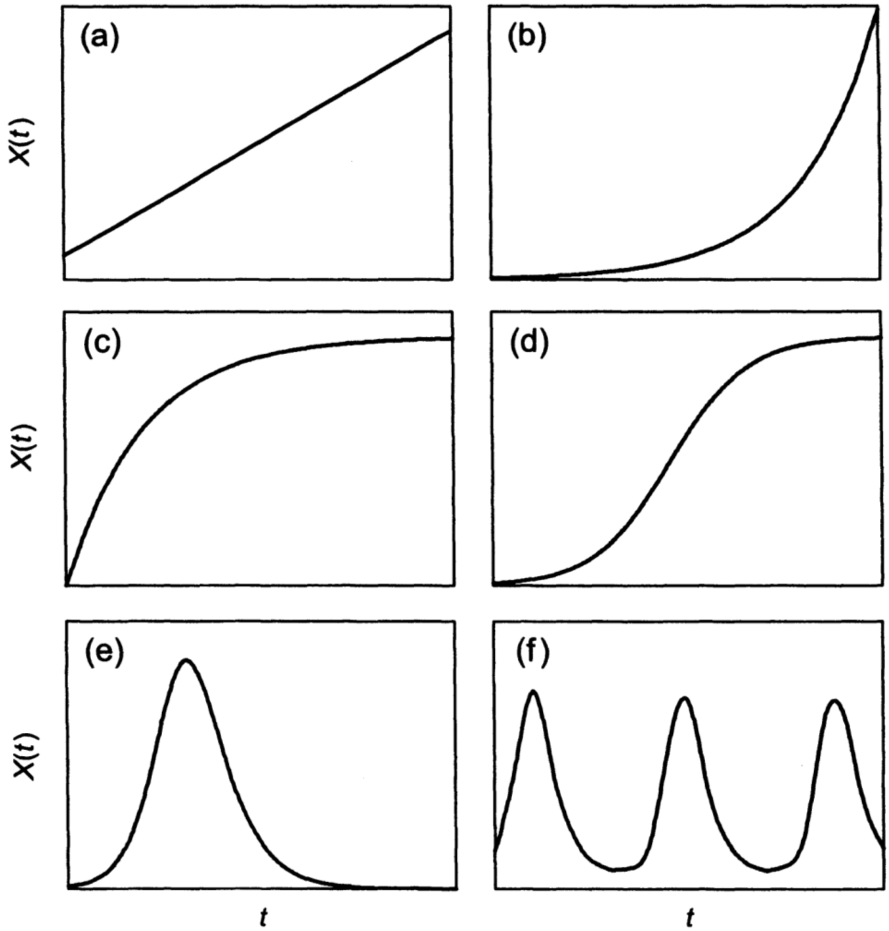

Figure 2.1 Qualitative types of dynamics: (a) linear, (b) exponential, (c) asymptotic, (d) logistic, (e) boom and bust, and (f) sustained oscillations.

Another simple model of growth obeys the following differential equation:

$$\dot{X} = rX \tag{2.2}$$

This is the *exponential growth* model. The parameter r is another constant, called the *relative rate of change* (because the total rate of change is the product of the relative rate r and X). Because the rate of change of X is assumed to be proportional to X itself, this type of growth is sometimes called *autocatalytic*: the more X there is, the faster it grows. The exponential equation provides the simplest model for the growth of biological populations and can be thought of as the first law of population dynamics (Turchin 2003). The autocatalytic part arises because the more animals there are in the population, the faster the population grows (since each animal can reproduce). The solution of the exponential model is a curve bending up (Figure 2.1b).

Copyrighted material

The linear and exponential models are examples of *boundless growth*. Such models often provide good starting points for modeling dynamical systems, because they make minimal assumptions about the system. In other words, they are *null models*, and that is why the first laws of mechanics and population dynamics belong to this class. But boundless growth models, by themselves, are not good models for the overwhelming majority of dynamic phenomena, because few real-life systems exhibit limitless growth. We need to add other mechanisms to the right-hand side of equations.

2.1.2 Equilibrial Dynamics

Few real-life processes grow without bound. Usually, there are some mechanisms—generally called *negative feedbacks*—that act to impose upper and lower limits on growth. One of the most important characteristics of a negative feedback mechanism is the lag time with which it operates. Some feedback mechanisms operate on a time scale that is much faster than the time scale at which the modeled variable X changes. In such cases we usually neglect the lag and assume that the feedback is instantaneous. A simple model that adds an instantaneous negative feedback to the linear growth is

$$\dot{X} = c - dX \tag{2.3}$$

Here two processes affect the dynamics of X. One force increases X at a constant rate c, but there is also a counteracting force, whose strength increases proportionately to X. At some point (specifically, when X reaches $X^* = c/d$), the strengths of the positive push and negative pull balance each other, and X stops growing. The point X^* where the rate of change of X is zero is called an *equilibrium*. Equilibria can be stable or unstable. For this model, the equilibrium X^* is stable, because when X is below the equilibrium, the positive push overwhelms the negative pull; and vice versa, if X somehow gets above X^*, then the negative pull will overwhelm the positive push, and return X back to the equilibrium. A typical trajectory for X predicted by equation (2.3) is initially linear (at low X) and then slows asymptotically as X approaches equilibrium (Figure 2.1c). I will refer to such dynamics as linear-asymptotic or *asymptotic growth*, for short.

Adding a negative feedback to the model of exponential growth is also simple. In this case, let us assume that the relative growth rate r is a linear function of X: $r(X) = r_0 - gX$. This leads to the *logistic* equation

$$\dot{X} = r(X)X = (r_0 - gX)X \tag{2.4}$$

Logistic growth is illustrated in Figure 2.1d.

Both asymptotic and logistic dynamics belong to the class of single-dimensional or first-order differential models. The general form of first-order models is $\dot{X} = f(X)$, where $f(X)$ is some arbitrary function of X. For example, in the logistic model, $f(X) = (r_0 - gX)X$ is a quadratic function of X. These models are called first order (or one dimensional) because there is only one structural variable, X. (In mathematical applications structural variables are called *state* variables, because they describe the state of the system. The use

Copyrighted material

of this standard terminology in dynamical systems, however, would create unnecessary confusion because the main subject of this book is states in the meaning of polities.)

In addition to the kinds of dynamics illustrated by the asymptotic and logistic models, in which the system is always attracted to a unique stable equilibrium, single-dimensional models can also have a more complex behavior, called *metastable* dynamics. For example, if $f(X)$ is a cubic polynomial, so that the model is

$$\dot{X} = a(X-b)(c-X)X \tag{2.5}$$

(a, b, and c are positive constants), then there are three equilibria: two stable ones (a low and a high equilibrium), and one unstable in the middle. If X is initially below the unstable equilibrium, the trajectory will be attracted to the low equilibrium. Alternatively, if X starts above the unstable point, the trajectory is attracted to the high equilibrium. One social science application of such an equation is in modeling "tipping" behaviors (see Chapter 6).

One very important fact that we need to know for later is that first-order differential models are incapable of oscillatory dynamics. They cannot even exhibit a single (rise-and-fall) oscillation. *Fast negative feedbacks, operating without an appreciable lag, cannot cause a dynamical system to oscillate.* They can only cause it to return asymptotically to an equilibrium, if a stable equilibrium exists (if it does not, then the system will run away to $\pm\infty$). In order for oscillations to arise, a negative feedback must operate with a delay. We model such slow feedbacks either explicitly, by adding other structural variables to the differential equation model, or implicitly by using discrete-time (difference) models.

2.1.3 Boom/Bust Dynamics and Sustained Oscillations

In order to model rise/fall dynamics of X with differential equations, we need to identify another structural variable, call it Y, which is affected by X and, in turn, itself affects X. Thus, X is affected by a negative feedback loop that is mediated by Y. Consider the following simple biological model of a population of consumers living on a nonrenewable resource:

$$\begin{aligned}\dot{X} &= caXY - dX \\ \dot{Y} &= -aXY\end{aligned} \tag{2.6}$$

where X is the number of consumers at time t and Y is the current amount of resources. Looking first at the Y equation, we see the term $-aXY$ representing the rate of consumption (the minus sign indicates that consumption reduces the amount of resources present). Consumption is assumed to be proportional to the product of X and Y, for the following reasons. First, more consumers deplete resources faster. Second, when resources are plentiful, each individual consumer depletes resources faster than when resources are scarce. Turning to the X equation, I assumed that consumers increase in proportion to the amount of resource consumed, with c being the constant of proportionality (this is the term $caXY$).

Copyrighted material

Additionally, consumers die off at a rate proportional to X (the proportionality constant d is known as the relative or per capita death rate).

The dynamics of this model are illustrated in Figure 2.1e. Assuming plentiful initial resources, consumers will first increase because their "birth rate" ($caXY$) will exceed the "death rate" (dX). Meanwhile, resources are depleted, and at an increasingly faster rate, because consumers are becoming more and more numerous. Eventually, resources fall beyond the point where consumer birth rate exceeds death rate, and the consumers start declining. Since consumers are still depleting resources, even during the decline phase, there is no end to the collapse: X will keep decreasing to 0. The boom is inevitably followed by the bust.

It is very easy to modify Model (2.6) to cause it to go through repeated boom/bust cycles. For example, we can add the assumption that the resource is renewable, and grows exponentially in the absence of consumers. Adding the exponential growth term bY to the second equation, we have

$$\dot{X} = caXY - dX$$
$$\dot{Y} = -aXY + bY \tag{2.7}$$

This is the famous Lotka-Volterra model of predator-prey cycles (a typical trajectory of X is shown in Figure 2.1f).

Second- and higher-order differential models (models with two or more structural variables) are capable of very diverse kinds of dynamic behaviors. They can have a stable equilibrium, approached either monotonically (as in one-dimensional models) or in an oscillatory fashion. They can exhibit stable cycles, characterized by a certain period and amplitude. Models with three or more structural variables can oscillate chaotically (predicting irregular-looking dynamics) or quasiperiodically (two or more cycle periods superimposed on each other). There are many kinds of fascinating mathematical phenomena, but we do not need to be concerned with them in the investigation of social dynamics, at least not for a long time yet. The important general class of dynamics for our purposes is second-order oscillations. Whether they are limit cycles, quasiperiodicity, or chaos is not critical at the current state of the art. The important feature, which distinguishes them from first-order dynamics, is sustained periods of increase followed by sustained periods of decline.

One further important result from nonlinear dynamic theory is that there is a close relationship between the time scales at which negative feedback loops operate (how fast they are) and the time scale of the dynamics (for example, the average period of oscillations). In differential equation models such as (2.7), the speed with which negative feedback acts is explicitly related to parameter values of the model, typically those whose units are [time] or $[\text{time}]^{-1}$. In the Lotka-Volterra model, there are two parameters whose units are $[\text{time}]^{-1}$: d and b. The parameter d measures how fast the numbers of consumers would (exponentially) decline in the absence of resource; b, analogously, measures how fast resources would (exponentially) increase in the absence of consumers. These two parameters determine the periodicity with which the consumer-resource system oscillates. In fact, the period is inversely related to the geometric mean of b and

Copyrighted material

d (for the mathematically inquisitive: the period of oscillations near the neutrally stable equilibrium is $2\pi/\sqrt{bd}$). Thus, the faster consumer and resource populations change in time, the shorter is the oscillation period. In models more complex than the Lotka-Volterra model the formula for the period is more complicated, but the qualitative insight carries over: faster feedbacks cause faster oscillations (and if feedbacks are too fast, then we cannot even obtain oscillations, because dynamics tend to be stabilized by very fast feedback mechanisms).

2.1.4 Implications for Historical Dynamics

Our discussion of dynamics, so far, has focused exclusively on *endogenous* factors—variables that participate in dynamic feedbacks. In a purely endogenous system any fluctuations are solely a result of the interaction of endogenous variables; such systems are "closed" with respect to influences from outside. Historical social systems, by contrast, should always be affected by outside forces: climate fluctuations causing crop failure, sudden appearance of new epidemics, hostile invasions, spread of new religions, and so on. Factors that influence a dynamical system, but are not themselves influenced by its variables, are called *exogenous*. The distinction between endogenous and exogenous factors is not sharp, and usually depends on the questions we choose. For example, if we are focused on the internal dynamics of a single polity, then we will model invasion by other polities as an exogenous factor. But if we decide to expand the model to cover the dynamics of the whole system of interacting states, or the *world-system* (Wallerstein 1974; Chase-Dunn and Hall 1997), then invasion occurrence is endogenized.

The open property of historical social systems presents no problem to the dynamical systems approach. The most natural way to model such influences is to add an exogenous structural variable to the system of equations. For example, if we already have endogenous variables X and Y, and add an exogenous variable Z, then the equations look something like

$$\dot{X} = f(X, Y, Z)$$
$$\dot{Y} = g(X, Y, Z)$$
$$\dot{Z} = h(t)$$

where f, g, and h, are some functions. That is, the rate of change of X and Y depends on all three variables in the system, while the rate of change of Z is given by some time-dependent function. There are two general ways to model Z: (1) as a stochastic variable or (2) as a deterministic trend. Which approach we choose depends on the nature of the exogenous variable, and the questions we wish to ask.

In Sections 2.1.1–2.1.3 I identified three fundamental types of dynamics characterizing purely endogenous systems. The simplest type encompasses systems that are not affected by negative feedbacks. I will call such dynamics *zero order*. Next, there are systems that are affected only by feedbacks acting very rapidly, which I will call *first-order* dynamics. Finally, there are systems that incorporate

Copyrighted material

multiple endogenous variables, leading to negative feedback loops acting with a time delay. These are *second-order* systems.

Allowing exogenous variables leads to a natural generalization of this order typology as follows. Zero-order systems are characterized by this general model:

$$\dot{X} = f(Z(t)) \tag{2.8}$$

In effect, X itself is the exogenous variable. Examples of such systems are various kinds of random walks (biased or otherwise), stochastic exponential increase or decline models, and so on. Such systems typically do not have an equilibrium density around which they fluctuate (unless we construct the function f in a very special way, to force such an "equilibrium" exogenously). Zero-order systems are not terribly interesting from the dynamical point of view, because any systematic dynamical patterns found in them are entirely due to the action of exogenous variables. The power of the dynamical systems approach would be largely misspent in applying it to such systems. However, zero-order dynamics provide a natural null model, against which other more complex alternatives can be tested.

First-order systems are governed by models of the form

$$\dot{X} = f(X, Z(t)) \tag{2.9}$$

where Z is an exogenous variable and does not depend on X. If Z is a stochastic variable, and f includes a negative feedback, then the dynamics are characterized by a stochastic equilibrium. X fluctuates in the vicinity of the stable equilibrium, and if X becomes too high or too low, endogenous dynamics push it back to the equilibrial level of fluctuations (in other words, the dynamic process is characterized by a return tendency). No cycles or any other kinds of complex dynamic behaviors occur in first-order systems, unless they are exogenously imposed (for example, Z oscillates periodically).

Second-order systems are governed by models like

$$\dot{X} = f(X, Y, Z(t))$$
$$\dot{Y} = g(X, Y, Z(t))$$

where Z is again an exogenous variable. More than two endogenous variables can be involved. Second-order systems are capable of all kinds of complex dynamics: stable equilibria, limit cycles, quasiperiodicity, chaos, multiple coexisting attractors, etc. Adding stochasticity expands the spectrum of possible behaviors even further. However, for the purposes of this book, I will call all such behaviors *second-order oscillations*. Perhaps the time will come when we have methods and data good enough for distinguishing between limit cycles and chaos in historical systems, but this time is not here yet. The fundamental importance of distinguishing between the three kinds of dynamics is that in zero-order systems all dynamics are driven exogenously, while in the first- and second-order ones some proportion of the variance in fluctuations is explained by the action of endogenous variables. Furthermore, different social mechanisms can often be classified as either fast or slow feedbacks, leading, correspondingly, to either first- or second-order dynamics. Detecting and characterizing such oscillation-inducing feedbacks is a major goal of the proposed research program in historical dynamics.

Copyrighted material

2.2 THE COLLINS THEORY OF GEOPOLITICS

One of the most powerful formulations of the geopolitical theory is found in the work of Randall Collins (1978, 1986, 1995); see Rozov (1997). Because Collins states his geopolitical principles clearly and succinctly, this verbal theory can be relatively easily translated into mathematical models. Furthermore, Collins and co-workers have also advanced formal geopolitical models, formulated as computer simulations (Hanneman et al. 1995). My plan in this section is to review the postulates advanced by Collins, and translate them into simple differential models. I will also do the same for the simulation model of Hanneman et al. (1995), thus bringing these developments to a common denominator. Another interesting model that is relevant to the issues at hand is the spatial simulation by Artzrouni and Komlos (1996).

2.2.1 Modeling Size and Distance Effects

As is natural, Collins' thinking has evolved over the last two decades, so the material below is based on his 1995 article, specifically on his Figure 1 there, which I redraw here as Figure 2.2. The main variable of interest is the state's territory size, or area. Temporal change in this variable occurs as a result of war success. The positive arrow from "war success" to "territory size" indicates that when the state is successful in war, it gains territory, while war failure implies territory loss. Territory size also positively affects "geopolitical resources" (more taxes and

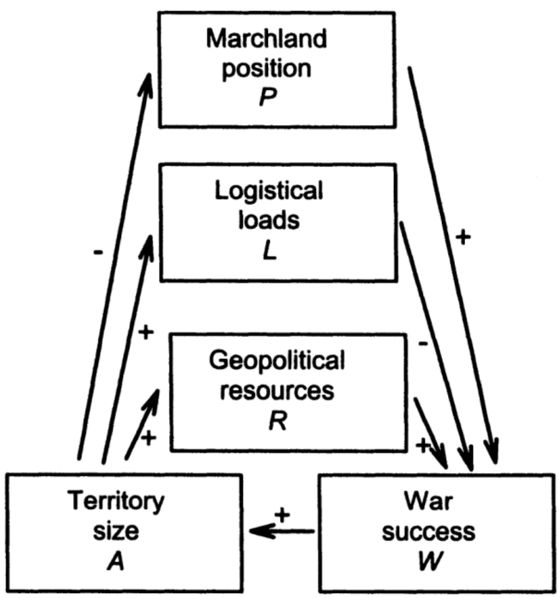

Figure 2.2 Feedback structure of the Collins geopolitical model. (After Collins 1995: Figure 1)

Copyrighted material

recruits for the army), which in turn positively affects war success. Increased territory size also means increased "logistical loads." The further that military power is projected from the home base, the higher the costs (Collins 1995:1558). Additionally, more state resources are tied up in policing the populace and extracting the resources. This connection between the state's size and logistical loads is often referred to as the "imperial overstretch" principle (Kennedy 1987). Increased logistic loads, in turn, have a negative effect on war success. Finally, "marchland position" favors war success, because states with enemies on fewer fronts expand at the expense of states surrounded by enemies (Collins 1995:1555). However, state expansion reduces the marchland advantage, as the state expands away from its initially more protected position.

We now translate this theory into formal models, starting with the loop involving geopolitical resources, and then adding the effect of logistical loads. Marchland position requires a spatially explicit approach and will be tackled in the next section. In developing the following models, I will make two general assumptions. First, and most importantly, I will assume that various feedbacks act rapidly with respect to territorial dynamics. Thus, I can use ordinary differential equations as the mathematical framework. Second, to make the model more concrete I will assume simple functional forms, usually linear ones. The effect of these assumptions on the results will be discussed below.

In the first model, there are three variables: territory size, or area A, geopolitical resources R, and war success W. I will assume that the rate of change of A is linearly related to war success: $\dot{A} = c_1 W$, where c_1 is a proportionality constant, translating war victory into square kilometers of territory gained. Resources should be roughly proportional to the area (in the simplest case, if population density is approximately constant, increased area implies greater population base to pay taxes and provide recruits). Thus, $R = c_2 A$. Finally, the relationship between resources and war success is a bit more complex. Resources translate into state power, but in order to gain victory, state power has to be greater than the power of the adversary. Assuming that the state we study (the *focal* state) exists in a homogeneous environment, characterized by a constant military power of rivals, we have $W = c_3 R - c_4$. The constant c_3 translates resources into power, while c_4 is the power of the adversary who must be defeated. The greater is the power of the focal state, in relation to the adversary power, the more successful it is in war, and, in consequence, the faster it increases its territory. Putting together these assumptions, and after some algebraic manipulations, we have the following model:

$$\dot{A} = cA - a \tag{2.10}$$

where I defined $c = c_1 c_2 c_3$ and $a = c_1 c_4$, to gid rid of unneeded parameter combinations.

The dynamics of this linear model are very simple, and depend on the initial territory of the state (see Figure 2.3a). If the initial A is below the threshold $A_0 = a/c$, then the rate of change is negative, and A decreases to 0. However, if A starts above the threshold, then it grows exponentially (that is, at an accelerating rate) to infinity. In other words, we are dealing here with a zero-order type of

Copyrighted material

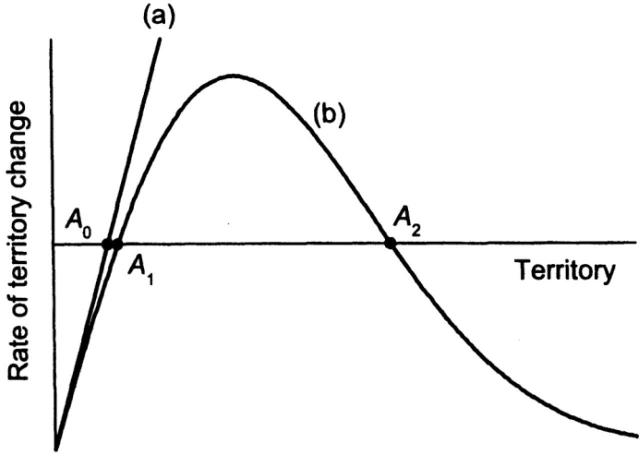

Figure 2.3 Relationship between the rate of territorial change, $\dot{A}$, and territory, A, in the two simple geopolitical models. (a) Model of territory size effects. (b) Model of territory size and distance.

dynamics. This is not at all surprising, because all arrows in the loop we modeled so far have pluses associated with them. In other words, we have just modeled a positive feedback loop, and obtained an entirely predictable result.

The loop involving logistic loads, on the other hand, involves one minus, and therefore it is a negative feedback loop. To model the effect of logistic load, let us follow Collins and assume that imperial overextension results from difficulties associated with projecting imperial power over distance. In other words, if state power at the center is P_0, then at distance r it is $P_0L(r)$, where $L(r)$ is the logistic distance multiplier, ranging from 1 at $r = 0$ to 0 at $r = \infty$. Boulding (1962:245–247) presents the argument that logistic distance multipliers should decline with increasing r as a negative exponential function: $L(r) = \exp[-r/c_5]$, with c_5 governing how rapidly power declines with distance. Since the relationship between area and radius is $A \sim r^2$, the logistic distance multiplier is $L(r) = \exp[-\sqrt{A}/c_7]$. Substituting this relationship in the model, we have

$$\dot{A} = cA\exp[-\sqrt{A}/h] - a \qquad (2.11)$$

where parameter combinations were again replaced with single parameters. Figure 2.3b shows that state power, as measured by its ability to expand, is negative at low A (just as in the pure size model), and initially increases with A. However, eventually the effect of increased logistic load begins to be felt. At $A = 4h^2$ state power is maximized, and for $A > 4h^2$ it begins to decline. There are two equilibrial points, A_1 and A_2. A_1, similarly to A_0 in the size-only model, is unstable: if the initial condition is below A_1, then the state is eaten by its neighbors. A_2 is a stable equilibrium: below it, the state's ability to expand is positive and A increases; above it, the state expansion rate is negative and A declines (Figure 2.3b).

Copyrighted material

of unit squares, with linear dimensions of about 40 km. At time = 0 (which was assumed to correspond to 500 C.E.), the simulated space was filled with equal-sized states, each occupying 5×5 squares, and thus with a starting area of roughly 40,000 km^2 (see Figure 2.4).

The two variables that affect the power of states in the Artzrouni and Komlos simulation are the area A (measured as the number of unit squares that make up a country) and the perimeter C (the number of squares that have a foreign neighbor). The power is increased by greater A, but decreased by greater C, because longer boundaries require more resources to defend. Note that this is a different conceptualization of logistic loads from the one used in the previous section. The marchland effect was modeled by treating state boundaries along the sea and mountains differently from those where no natural borders are present. Artzrouni and Komlos assumed that a sea boundary is easier to defend. Therefore, in calculating the boundary length, each unit along the sea counted only as a fraction f of a unit along the land border. In addition, two European mountain chains, the Pyrenees and the Alps, were assumed to present impregnable natural barriers, so that unit squares abutting these areas were not counted as part of C. However, the Pyrenees and the Alps did not completely cut off their respective peninsulas from the continent (see Figure 2.4).

Artzrouni and Komlos assumed the following specific form for the power function of state i:

$$P_i = \frac{A_i}{\alpha + \exp[\gamma C_i + \beta]} \tag{2.13}$$

where A_i and C_i are the area and the perimeter of state i, respectively, and α, β, and γ are positive constants. This function, although conceptualized differently, results in the same general shape of the relationship between the state size and power as in Model (2.11). As we increase A from 0, P first increases with A, as a result of the positive feedback associated with increased geopolitical resources. Eventually, however, the negative feedback associated with increased logistic loads overpowers the positive one, and for very high A, P declines to 0.

The relative powers of states determine their success in war with neighbors. Each iteration of the model is made of one bilateral interaction (war). The simulation model chooses a country at random (call it i) and compares its power to the powers of all its neighbors. The simulation then determines which neighbor j differs the most in power from i (the j for which the absolute difference $|P_i - P_j|$ is maximized), and the two countries go to war. The more powerful country wins with probability $1 - 0.5\exp[-K(P_i/P_j - 1)]$; otherwise, the weaker country wins. After the conclusion of war, all boundary squares of the loser contiguous to the winner are absorbed by the winner. The simulation then performs another iteration, choosing a country at random, etc. Each iteration corresponds to 1/3 year.

The Artzrouni and Komlos model has five parameters (f, α, β, γ, and K). The investigators used the method of trial and error to find the specific values of these parameters that would replicate the historically observed dynamic map of Europe as closely as possible. One particular realization of the simulation is shown in

Copyrighted material

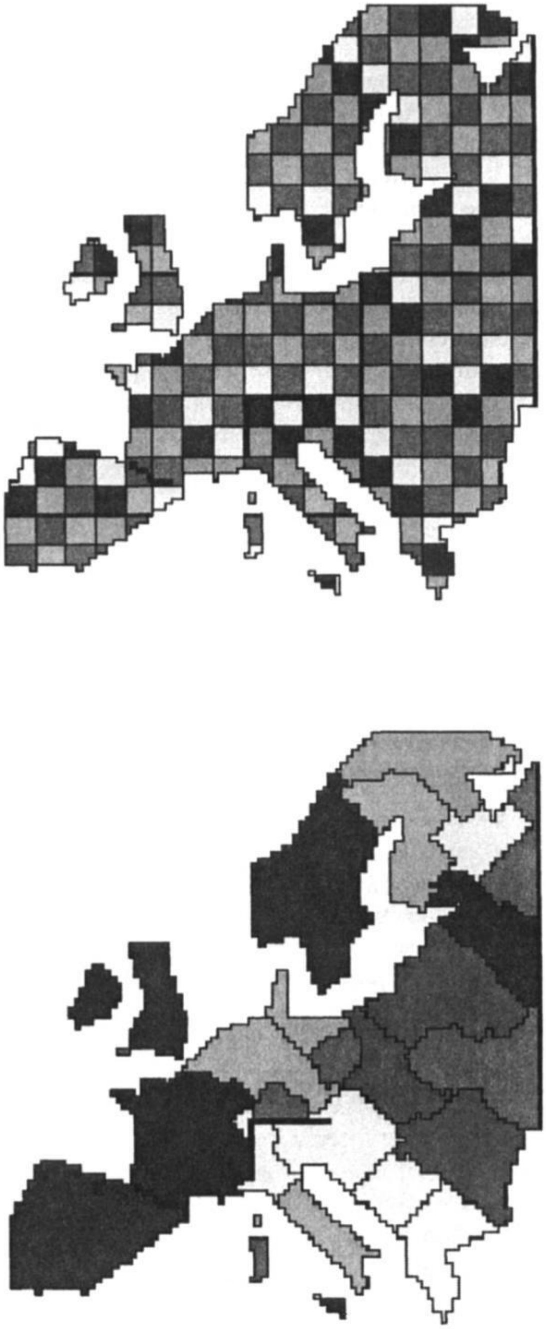

Figure 2.4 Territorial dynamics of the European state system as simulated by the model of Artzrouni and Komlos (1996). Top: initial map at 500 C.E. Bottom: model-predicted map at 1800 C.E. Thick black lines indicate the location of the Pyrenees and the Alps. (After Artzrouni and Komlos 1996: Figure 4)

Copyrighted material

Figure 2.4: the 234 initial countries at 500 C.E. are reduced to 25 by the end of the simulation. The outlines of coastal countries ("France," "Spain," and "Italy") take shape rapidly, while inland countries take longer to solidify. Additionally, there is much more variability in the final configuration achieved away from the coasts in different realizations. Thus, the marchland effect has two aspects. First, countries enjoying it achieve somewhat larger size compared to more centrally located ones (this can be seen by the large size achieved by "Spain," "France," and "Sweden/Norway" in Figure 2.4). Second, their boundaries reach stability much faster than the boundaries of inland states.

To summarize, the Artzrouni and Komlos simulation provides a confirmation of the postulated effect of the marchland advantage. Additionally, the simulation suggests that the boundaries of present states, especially those with long sea borders (Spain, France, Italy, and Greece), may be determined in a large degree by geopolitical mechanisms. However, Artzrouni and Komlos are very careful to stress that simulation parameters must be tuned just right for the simulation to reach the desired equilibrium. While the circularity involved in parameter calibration weakens the result, we should keep in mind that the simulation is very parsimonious, having only five free parameters. The sensitivity to one parameter, f, is of particular interest to the question of the marchland effect. If f is set too low, then the simulations usually yielded just one or two countries with maritime borders to the west of Europe that swept eastward across the continent. Their small effective perimeters kept their power high, allowing them eventually to annex all their neighbors.

Finally, the geopolitical simulation of Artzrouni and Komlos, similarly to analytical models advanced before, generates an essentially first-order behavior. Large countries conquer smaller and eventually expand until they reach the limits set by imperial overstretch, where their size is stabilized. In other words, the model does not predict that states reaching too high a size will collapse.

2.2.3 Conflict-legitimacy Dynamics

The geopolitical theory of Collins has three major principles (Collins 1986:168, 1995: Figure 1): (1) territorial resource advantage, (2) marchland advantage, and (3) overextension as a result of increased logistic loads. Our modeling in previous sections suggests that these postulates lead to first-order dynamics characterized by initially accelerating territorial growth that eventually reaches a stable equilibrium. Geopolitical models do not predict the collapse of large powerful empires (although small states may be destroyed before they manage to grow above a critical size). Yet, historical empires exhibit a different behavior, because they always eventually collapse. When a dynamical system exhibits opposite trends (growth versus decline) for the same values of variables in the explanatory set, this means that there is another hidden variable that determines the direction of change, which we have not yet included in the set. Thus, our modeling efforts have already paid for themselves: they showed that we need to look for explanatory mechanisms other than pure geopolitical principles, in order to understand the rise and collapse of territorial empires. It appears that Collins

Copyrighted material

has also reached the same conclusion, judging by his extensive discussion of mechanisms of state breakdown in the 1995 paper. Specifically, Collins discusses two theories: (1) the demographic-structural model (Goldstone 1991b), and (2) ruler legitimacy as affected by geopolitical power-prestige (Hanneman et al. 1995). Dynamical systems theory suggests that the key property of the postulated mechanisms of collapse is the time scale on which they act (the concept of temporal scale is also discussed by Collins; e.g., Collins 1995: Figure 6). Thus, our task, which will be largely pursued beyond the confines of this chapter, is to translate various postulated nongeopolitical mechanisms into models, determine whether these models are in principle capable of generating second-order dynamics, and, if so, derive testable predictions from them.

Although the conflict-legitimacy model of Hanneman et al. is not based on a geopolitical mechanism (in the strict sense), I will review it in this chapter, because this model is already well developed (and can be quickly summarized) and is closely connected with the models considered earlier. Actually, Hanneman et al. develop not one model, but a series of models of increasing complexity. This is a methodologically sound approach; in fact, I am in complete agreement with the philosophy of modeling as set out by Hanneman et al. To their excellent recommendations (see also Hanneman 1988), I would add only that more attention should be paid to a parallel development and consideration of analytical models.

The core of the theory advanced by Hanneman et al. focuses on the interaction between power-prestige, state legitimacy, and international conflict. Hanneman et al. assume that the motivation of rulers to initiate external conflict is directly proportional to the difference between their current legitimacy and the goal of maximum legitimacy. For any given level of conflict initiated, the degree of success is determined by the proportional superiority of the power of the focal state, relative to that of its rivals. Change in the state prestige is proportional to war success, and legitimacy follows, with delay, from prestige (Hanneman et al. 1995: 17). Hanneman et al. do not explain why legitimacy should follow war success with a lag time. Yet, in their model they impose a substantial lag: whereas war occurs at every time step (they use a discrete-time framework), legitimacy follows with a lag of three time units. This would suggest that (crudely) on average it takes victory in three successive wars for the legitimacy of the state to increase substantially. I would argue, by contrast, that legitimacy operates on a much faster time scale. War victory is immediately followed by a rush of patriotism that floats up the fortunes of politicians, and, vice versa, war failure is immediately followed by disillusionment with the powers that be. If legitimacy were a slow variable, then there would be much less temptation for politicians to use "a small victorious war" to bolster their shaky legitimacy.

Hanneman et al. develop three models: (1) the core model that focuses on war legitimacy dynamics, (2) a more complex version that adds the costs and benefits of empire, and (3) the full model that further adds imperial capitalism and the military-industrial complex. Of particular interest to us is the second model because, for certain parameter values, it appears to predict repeated instances of imperial growth followed by breakdown. However, it appears that occurrence of these instances of state collapse depends in a critical way on the assumption of a

Copyrighted material

lag time with which legitimacy follows war success. To check whether this is true, I translated the core of the Hanneman model into a differential equation model. In the Appendix (see Section A.1) I show that *the legitimacy-conflict model is described by a single-dimensional ordinary differential equation*. In other words, we again end up with a first-order model. This model can have multiple equilibria, and depending on the initial values of territory size and previous record of war success, the trajectory will be attracted to one or another of the stable ones. But the model is incapable of exhibiting boom/bust dynamics or sustained oscillations. The inescapable conclusion, therefore, is that the interaction between legitimacy, war success, and territorial expansion cannot generate sustained imperial decline. Thus, the imperial collapses occurring in the Hanneman model appear to be entirely due to the assumed delay with which legitimacy follows war success.

2.3 CONCLUSION: GEOPOLITICS AS A FIRST-ORDER PROCESS

I started this chapter by reviewing some elementary facts from nonlinear dynamics for the following reasons. Most social scientists are not closely familiar with dynamical systems theory, and I wanted to present a nontechnical summary of its insights most relevant to the issues dealt with in this book. One such particular insight is that there is a close relationship between the time scales at which negative feedbacks operate and the nature of the dynamics. If feedback mechanisms operate much faster than the dynamics of the focal variable, then the system cannot oscillate or even exhibit a single boom/bust behavior. If we do have an oscillatory system, then, more quantitatively, the speed with which a feedback acts determines the temporal pattern of the dynamics (for example, the average length of an oscillation, or a boom/bust cycle). This means that if a feedback loop operates on the scale of years, or even worse, weeks, then it is highly unlikely that it could cause oscillations, whose average period is measured in centuries. Centuries-long cycles are typically caused by feedbacks operating on the scale of human generations (decades or longer).

This insight is very relevant to the issue of what mechanisms underlie imperial boom/bust cycles. Empires grow and decline on the time scale of centuries (Taagepera 1978a, 1978b, 1997; see also Figure 4.4 in Chapter 4). Let us make a simple analysis of the imperial growth/decline data tabulated in the Appendix of Taagepera (1997). Taagepera defined the rise phase as the time it takes for a polity to expand from 20% to 80% of its maximum area (1997: 480). We can define the decline phase analogously, as the time needed to decline from 80% to 20% of the maximum, and the peak phase as the time from the end of the rise to the beginning of the decline. Table 2.1 gives the phase durations for the 31 polities from Taagepera (1997), that had four or more consecutive area observations (we need these data points to unambiguously define the phases). There is a large amount of variation in the durations of decline phases for these polities. In about half of the cases (14) the decline phase was on the order of one human generation (less than 0.3 centuries). The rest of empires exhibited longer decline phases,

Copyrighted material

act on a fast scale. First-order dynamics are equilibrial; examples include asymptotic and logistic growth processes. First-order dynamics may also be metastable (more than one stable equilibrium is present). Finally, second-order dynamics arise in systems in which dynamical feedbacks act with a lag. Examples of second-order behaviors include a single boom/bust dynamic and sustained periodic or chaotic oscillations.

- The geopolitical theory of Randall Collins postulates three main mechanisms explaining territorial dynamics of states: geopolitical resources, logistical loads, and the marchland position.
- The mathematical model incorporating only the positive feedback between territory and geopolitical resources exhibits zero-order dynamics. If the initial state territory is above a certain threshold, then it grows in an accelerating fashion. However, if the initial territory is below the threshold, then the state shrinks and eventually disappears.
- Adding to the model the negative feedback of the logistical loads leads to first-order dynamical behavior, metastability. Again, if the initial territory is below the threshold, the state loses ground and disappears. However, starting above the threshold, the territory does not increase without bound, as in the simpler model, but approaches an upper equilibrium. This equilibrium is stable with respect to small perturbations.
- In order to examine the positional effects I turn to a spatial simulation model developed by Artzrouni and Komlos. This model suggests that states initially enjoying marchland advantage (a higher proportion of boundary along a coastline) grow to larger sizes than inland states. However, the model does not exhibit any second-order oscillations: the loser states disappear, while the winners grow to the limits set by logistical factors, where their size is stabilized.
- Finally, I review the simulation model of conflict legitimacy dynamics developed by Hanneman and co-workers. I show that if we translate this model into differential equations, then we again obtain a first-order system that is incapable of second-order oscillations.
- An analysis of growth/decline data tabulated by Rein Taagepera suggests that long periods of imperial decline (more than a century) are frequently found in the historical record (12 cases out of 31). This finding strongly suggests that at least in some historical cases imperial dynamics were governed by second-order mechanisms. However, models based on purely geopolitical mechanisms do not predict such prolonged declines. Thus, we must investigate other mechanisms of imperial collapse.

Copyrighted material

Chapter Three

Collective Solidarity

3.1 GROUPS IN SOCIOLOGY

3.1.1 Groups as Analytical Units

In the previous chapter I suggested that we cannot understand the territorial dynamics of polities without studying their inner workings. This raises an important question: what are the elementary units in terms of which our theories should be constructed? The philosophical principle of methodological individualism maintains that ultimately sociological theories should be based on the properties of individuals. I agree with this approach in principle, especially if we stress the key word *ultimately*. However, methodological individualism, in my opinion, must be tempered with two important caveats. First, the idea that individuals are somehow more "real" than groups does not appear to be tenable. Human individuals cannot exist apart from a group and remain human (as real-life "Mowglis" attest). Furthermore, human groups are more than simple collections of individuals. Unlike animal groups, human groups are uniquely able to plan and purposefully carry out actions (Alexander and Borgia 1978; Melotti 1987).

Second, an attempt to follow the prescription of methodological individualism in one step does not appear to be a good modeling strategy (Section 1.2.3). Polities, especially such complex ones as empires, contain multitudes of individuals differing among themselves in a multitude of ways. Furthermore, an individual primarily interacts with a small subset of others, rather than directly with everybody else in the polity. In other words, large human societies consist of a number of groups, often hierarchically nested within each other. Thus, a much better modeling strategy would be to break the problem into two (or more) steps. First, we would like to understand how group dynamics arise from individual action, and then we can use group properties to model polity dynamics. "There is a distinctly sociological way of looking at the world. It holds that the key to understanding social life lies with the analysis of groups, rather than individuals." (Hechter 1987:2) An excellent example of such a hierarchical approach that introduces groups as intermediate actors between individuals and social dynamics is Jack Goldstone's (1994) analysis of revolutionary action. We also should keep in mind that eventually it will be necessary to progress to the next level and consider how polities interact within world-systems.

There are two characteristics that are particularly responsible for making human groups not just collections of individuals, but agents in their own right: the tendency to draw social boundaries and the capacity for group-oriented action even if it is individually costly.

Copyrighted material

Social Boundaries

Humans use many cues to demarcate group membership (Shaw and Wong 1989; Masters 1998). One of the most important recognition markers is language, especially dialect, accent, and speech patterns. For example, there is abundant experimental evidence from several societies that people are more disposed to cooperate with others who have the same dialect as themselves, even when dialectal differences are slight (Nettle 1999:57). Phenotypic similarity provides a number of potential markers: visible resemblance of facial and body form (and even odor); movement patterns, facial expression, and behavioral stereotypes; clothing and ornaments; and style and manners (the latter are especially important for signaling social class). Speech dialect and phenotype provide obvious, instantaneous information about group membership. Other categories of markers include kinship (presumed common descent; can be fictitious), religion (shared beliefs, norms, and rituals), and territory or proximity of residence (Masters 1998:456–457). The last category can also include shared membership in the same polity (nationalism or"regnalism"; see Reynolds 1997).

Capacity for Solidaristic Behaviors

A very powerful approach for inferring patterns of collective action from individual behaviors is the rational choice theory (Coleman 1990). The basic premise of this theory is that individuals are utility maximizers. The rational choice theory, however, has been unable to solve one very important problem in sociology: how societies can function without falling apart. An important theory (as formulated, for example, by Thomas Hobbes) maintains that society is based on the concept of social contract. However, it turns out that if people acted on a purely rational basis, they would never be able to get together to form society at all (Collins 1992:9). In fact, this "nonobvious sociological insight" (Collins 1992) approaches, in my view, the logical status of a theorem. For example, Kraus (1993) shows how the best developed theories of Hobbesian contractarianism all founder on the collective action problem, the free-rider predicament (Olson 1965; for a nontechnical review, see Collins 1992:13–19).

There appears to be only one solution to the puzzle of how societies can hold together (Collins 1992). Although people pursue their selfish interests most of the time, they also have feelings of solidarity with at least some other people. Such *precontractual solidarity*, in Durkheim's words, is the basis of societies (Collins 1992). States and armies break apart when people stop thinking of themselves as members of the group and think only of their own individual self-interest (Collins 1992:23).

Thus, the functioning of society can only be understood as a mixture of self-centered (rational) and group-centered ("extrarational") behaviors. In Collins' (1992:8) view, "the overall structure of society is best understood as a result of conflicting groups, some of which dominate the others. But conflict and domination themselves are possible only because groups are integrated at the micro level." This statement captures very nicely the essence of the approach that I develop here.

Copyrighted material