Natural Experiments of History

EDITED BY

Jared Diamond
James A. Robinson

THE BELKNAP PRESS OF HARVARD UNIVERSITY PRESS
Cambridge, Massachusetts • London, England • 2010

Historical Review 72 (1967): 469–496; François Furet, *Interpreting the French Revolution* (New York, 1981).

61. Comparative evidence on agricultural productivity is presented in Allen, *Economic Structure*, and on real wages in Robert C. Allen, "The Great Divergence in European Wages and Prices from the Middle Ages to the First World War," *Explorations in Economic History* 38 (2001): 411–447. Jean-Laurent Rosenthal, *The Fruits of Revolution: Property, Litigation and French Agriculture, 1700–1860* (New York, 1992); Philip Hoffman, "France: Early Modern Period," in Joel Mokyr, ed., *The Oxford Encyclopedia of Economic History* (New York, 2003), argues that the current consensus among economic historians is that the institutions of ancien régime France were indeed to some extent responsible for its relative underdevelopment.
62. Rondo E. Cameron, *France and the Economic Development of Europe, 1800–1914: Conquests of Peace and Seeds of War* (Princeton, NJ, 1961).
63. David S. Landes, *The Unbound Prometheus* (New York, 1969), pp. 142–147.
64. Timothy C. W. Blanning, "The French Revolution and Modernization in Germany," *Central European History* 22 (1989): 109–129; Hamerow, *Restoration, Revolution, Reaction*, pp. 22, 44–45.
65. Kisch, *Domestic Manufacture*, p. 212.
66. Jeffry Diefendorf, *Businessmen and Politics in the Rhineland, 1789–1834* (Princeton, NJ, 1980), p. 115.
67. Blanning, *The French Revolution in Germany*.
68. The analysis we present here is a simple exposition of the basic finding of Acemoglu et al., "The Consequences of Radical Reform," where a full statistical analysis is presented of the impact of French reforms on urbanization in Germany.
69. Although Figures 7.1 and 7.2 suggest that urbanization may have been growing more rapidly in either treatment group before 1800, which suggests that these areas had started to do differentially well even before the French invasion, in Acemoglu et al., "The Consequences of Radical Reform," we show that this difference is not statistically significant and in neither cross-national nor within-Germany data is there evidence of differential economic progress prior to the French Revolutionary period.

Afterword: Using Comparative Methods in Studies of Human History

JARED DIAMOND AND
JAMES A. ROBINSON

All natural experiments challenge scholars with certain recurrent types of methodological problems.[1] Similar challenges also present themselves, to varying degrees, in manipulative laboratory experiments and in the physical and biological sciences. For instance, no two human systems differ solely with respect to the single variable whose effects interest the scholar. Instead, there are inevitably other differences, which may also contribute to or dominate the outcome that one is measuring. No magic bullet or formula has been discovered for solving these and other challenges raised by natural experiments, just as no formula has been devised for solving the challenges involved in writing narrative history or in doing manipulative experiments. However, we can offer some suggestions. At minimum, it will help to be alert to these problems, and to learn from other scholars who have wrestled with them, as illustrated by the authors of the chapters of this book.

Let us start with a classification of natural experiments. One can think of them as involving differences either in perturbations or in initial conditions. Of course, this distinction is oversimplified, for reasons that we shall discuss below. The table on pages 258–259 lists the main differences of each type operating in the eight studies of this book.

In one type of natural experiment, different outcomes result especially from variation in the perturbation; the differences in the

The eight case studies of this book

(1) Chapter	(2) Subject	(3) Number of cases compared	(4) Initial conditions	(5) Perturbation	(6) Outcome examined
1	Polynesian cultural evolution	3	*Islands with different physical environments*	Polynesian settlement (+ different durations of settlement)	Sociopolitical and economic complexity
2	Frontier societies	7	Different temperate non-European lands	Explosive frontier settlement (from different sources at different times)	Cycles of boom, bust, and export rescue
3	New World banking systems	3	*Different political institutions, wealth, and income distributions*	Need for banks	Form of banking system
4a	Hispaniola	2	Two halves of the same island (different rainfalls, slopes, and soils)	*French vs. Spanish colonization* (+ different dictators)	Wealth, export economy, forest cover, and erosion
4b	Pacific islands	81	*Islands with different physical environments*	Human settlement (+ different colonizing peoples and durations of settlement)	Deforestation
5	Africa's slave trades	52	Different parts of Africa (+ different physical environments, resources, religions, precolonial development, colonizing powers, and legal systems)	*Slave trade present or absent* (+ 4 different slave trades)	Current income
6	Public goods in India	233	Different parts of India (+ different pre-colonial agricultural productivity, religion, and population composition)	*3 different colonial land tenure systems*	Schools, electricity, and roads (+ literacy, electoral competition, and outcome)
7	Effects of the French Revolution	29	Different parts of Germany (+ different religions, prior rates of urbanization)	*3 different courses of Napoleonic invasion*	Urbanization, as a measure of economic growth

Note: This table characterizes the eight case studies presented in this book. Column 3 gives the number of cases compared in each study (e.g., number of islands, countries, or districts compared). Column 6 identifies the outcome to be explained in each study. Italicized items in columns 4 and 5 identify either the different initial conditions (e.g., different island physical environments, different political institutions: Chapters 1, 3, 4b) or different perturbations or presence or absence of a perturbation (e.g., slave trade or Napoleonic conquest present or absent: Chapters 4a, 5, 6, 7) mainly responsible for those different outcomes. Items in parentheses in columns 4 and 5 are potential explanatory factors that proved not important or less important for those outcomes. See text for discussion.

initial conditions (i.e., in the place or society to which the perturbation is applied) are of less significance to the outcome. Perturbations (referred to as "treatments" in much of the literature about experiments) can in turn be either "exogenous" or "endogenous," and the experiment can consist either of comparing a perturbation with no perturbation or of comparing different types of perturbations. The comparison of a perturbation with a nonperturbation is exemplified by areas of Africa subjected to slave trading or not (Chapter 5), and areas of Germany invaded by French Napoleonic armies or not (Chapter 7). Different types of perturbation are the two halves of Hispaniola colonized by Spain or France (Chapter 4), three different systems of land tenure and revenue imposed on different parts of India by Britain (Chapter 6), institutions imposed by French conquerors on Germany being subsequently left intact or else reversed (Chapter 7), and settlement of non-European frontiers originating from four different European countries (especially salient being whether English or not) and at different times during the Industrial Revolution (Chapter 2). All of these perturbations can be considered exogenous—that is, arising from outside the studied area.

In the other type of natural experiment, the perturbation is similar across all cases, and the different outcomes instead result mainly from differences in the initial conditions. We have two examples from Pacific islands differing greatly in physical environment (especially in area, elevation, isolation, geology, and climate), all settled by a single colonizing people (Polynesians: Chapter 1) or else by Polynesians and related groups of Pacific island peoples (termed Melanesians and Micronesians: Chapter 4), with the outcome examined being either socioeconomic and political complexity (Chapter 1) or deforestation (Chapter 4). Our third example is of three New World countries differing greatly in their political institutions, as well as in their wealth and their income equality (Chapter 3). In that study, the "perturbation" can be thought of as an endogenous one: the common need for a banking system arising within

countries originally lacking chartered banks, as opposed to the exogenous perturbations such as invasions and imposed revenue systems examined in our other studies.

Most of our case studies focus on explaining *differences* in outcomes related to different perturbations or to different initial conditions. Of equal interest, however, are cases in which similar outcomes emerged despite big differences in the perturbations or initial conditions. The most striking conclusion of Chapter 2, which compares the development of frontier settler societies in seven former European colonies, is the similarity in outcomes. Despite differences in the "perturbations," especially in the European sources of immigrants and institutions, and in the stage of the Industrial Revolution at which these frontiers exploded, one is struck by similarities between these frontier societies—especially in their three-stage cycles of booms, busts, and export rescues, but also in many other respects including their growth of cities, transport infrastructure, wood consumption, farms and farm animals, their shared problems of infusing capital and immigrants and imports, their conquest of distance as a prerequisite to their export booms, their impacts on indigenous peoples, and their changing attitudes toward immigrants. These shared features evidently arose from the similar internal dynamics of growth in all of those frontier societies, overriding their differences in European sources of immigrants and institutions and in their decades of explosion. At the same time, the similarities in outcome were accompanied by differences in outcomes—for example, in the percentage of the immigrants who returned to their European mother countries, and in the frequency and duration of boom/bust/rescue cycles.

Our distinction between initial conditions and perturbations is not entirely sharp. Although there is no doubt that different sizes of Pacific islands constituted differing initial conditions to Polynesian settlers, and that invading Napoleonic armies (or their absence) constituted a perturbation to German principalities, how should one characterize the different political institutions and wealth of nineteenth-century Brazil, Mexico, and the United States for purposes of understanding their banking systems? Those differences constituted initial conditions insofar as they already existed before any of those countries had chartered banks, but institutions and wealth changed during the nineteenth century, and the banking systems may have been contributing causes as well as outcomes of the differences in wealth.

We have chosen, for illustrative reasons, to focus on case studies in which different outcomes can be attributed mainly either to differences in perturbations or to differences in initial conditions. However, one can also compare cases differing simultaneously in perturbations and in initial conditions, and the importance and interest of those cases may make the comparison profitable despite the added complication of having to consider both types of differences.

A question inevitably arising in any comparative study that compares perturbed societies or sites with nonperturbed ones concerns the perturbers' "selection" of which particular sites to perturb. In a laboratory experiment comparing so-called experimental and control test tubes that are identical except for some perturbation effected by the experimenter (e.g., adding one chemical to some but not other test tubes), the selection of experimental and control tubes can indeed be made completely random with respect to the experimenter's decisions. For example, the experimental-versus-control status of each test tube can be determined by flipping a coin or by using a random-number generator. However, important historical decisions are rarely made by flipping coins: Napoleon did have his reasons for invading certain German principalities but not others (Chapter 7), just as slave traders had their own reasons for buying slaves from certain parts of Africa but not others (Chapter 5). Thus, the practical question that the comparative historian must always ask is: were the perturbed sites selected for reasons irrelevant to the outcome studied (i.e., "random" with the respect to that outcome)? Or were the perturbed sites selected on the basis of differences in initial conditions material to the outcome?

All of our case studies comparing perturbed to nonperturbed sites, or comparing sites exposed to different types of perturbations, explicitly address this question and amass evidence showing that the grounds on which human historical actors selected particular sites for particular perturbations (or for lack of perturbations) do not explain the particular types of outcome studied. For instance, the analysis of Chapter 7 shows that areas of Germany invaded by French Revolutionary armies between 1792 and 1815 became more urbanized after 1860, but that is not because Napoleon preferred to invade already urbanized areas or because he presciently invaded areas likely to become more urbanized fifty years later. Instead, he chose his targets for contemporary military or dynastic or geopolitical reasons. His targets were actually on the average less urbanized at the time of his invasions than were the German areas that he spared. Similarly, British colonial administrators variously imposed three different land revenue systems in a geographic patchwork across India, and the analysis of Chapter 6 shows that one of those three types of patches (the patches with so-called landlord tenure systems) ended up more developed today by various indices. However, the type of revenue system imposed on each patch depended on either the colonial ideology that happened to be prevailing in Britain at the time that Britain annexed that particular patch, or else on the preference of the particular colonial administrator in power at that time, rather than on the contemporary development or other features of that patch relevant to development. This concern about patch selection should never be lightly dismissed. Instead, it must always be carefully evaluated in comparative studies in which perturbation is a variable—as opposed to studies in which all patches are more or less uniformly perturbed (e.g., by Polynesian settlement) but differ in initial conditions. Indeed, Chapters 5, 6, and 7 all use statistical techniques, especially instrumental variables regression, to investigate directly whether or not the perturbation is subject to problems of selection relevant to the outcome studied.

Historians seeking causal explanations would be fortunate if effective perturbations were followed promptly by their outcomes. In actuality, the outcome may be delayed by decades or even by centuries (e.g., if the perturbation alters societal or political institutions but those altered institutions do not produce the outcome under study until other changes accumulate).

For instance, western Hispaniola (Haiti) is today far poorer than eastern Hispaniola (the Dominican Republic), largely because of consequences of their different colonial histories (Chapter 4): France's colonization of the west ending in 1804 and Spain's colonization of the east ending initially in 1821. However, those different histories resulted in ex-French Haiti being much *richer* than the ex-Spanish Dominican Republic at that time of independence, and it took a century or more for the slowly developing consequences of those different colonial histories to result in the Dominican Republic overtaking and then far outstripping Haiti economically.

Again, the new institutions established in French-conquered areas of Germany before 1814 did not by themselves make those areas more urbanized and economically developed. Instead, the new institutions were more conducive to the Industrial Revolution (which is what brought urbanization and economic development) than were the old institutions swept away in conquered areas by Napoleon, but the Industrial Revolution did not begin to pay off in Germany until several decades after 1814.

Yet another possible example comes from the long-standing debate about why Europe eventually overtook China's earlier lead in technology, economic development, living standards, and power.[2] By many indicators, Europe began to pull ahead of China only in the 1700s and especially in the 1800s. Hence some authors seek explanations in causes emerging within those centuries themselves, such as Europe's Industrial Revolution and the trans-Atlantic trade. However, other authors see the fundamental causes much earlier, in medieval Europe's institutional development and agriculture or in much older European and Chinese geographic factors, which resulted in technological and economic growth only when industrialization and trade were added many centuries later. Such phenomena—which may be represented as "A + B together cause C, but only when B arrives long after A"—are as pervasive a problem for historians seeking to understand history as they are for psychologists and biographers seeking to understand individual human lives.

A ubiquitous concern in natural experiments is whether the different outcomes observed really were due to the particular types of differences in perturbation or initial conditions noted by the "experimenter," or whether they were instead due to some other difference. This risk of misinterpretation arises even in controlled laboratory experiments. A famous example was the discovery of the Josephson Effect in physics: laboratory measurements of superconductivity initially yielded confusing results, until Brian Josephson realized that a driving independent variable was slight temperature differences, to which superconductivity proved to be far more sensitive than had been previously realized. But this risk of misinterpretation due to variables other than those initially of interest is much greater in natural experiments, where one's variables are uncontrolled.

The natural experimenter should at least attempt to minimize the effects of individual variables other than those of interest, by choosing for comparison systems that are as similar as possible in other respects. For instance, in Chapter 7 of this volume, Acemoglu et al. restrict their comparisons of areas of Europe conquered or not conquered by Napoleon to German areas, in order to reduce cultural variation extraneous to the purpose of their study. However, in other related studies not presented in this book, Acemoglu et al. relaxed that restriction, examined non-German areas as well, and reached similar conclusions about Napoleon's effects. Kirch (Chapter 1) restricts his comparisons of Pacific island sociopolitical and economic complexity to islands colonized by Polynesians. However, in Chapter 4 Diamond relaxes this constraint by comparing Pacific islands colonized by Micronesians and Melanesians as well as by Polynesians, in order to examine an outcome variable (deforestation) that is expected to be less sensitive to differences among colonizing peoples than is the outcome of sociopolitical and economic complexity studied by Kirch. Diamond compares the two halves of the Caribbean island of Hispaniola differing in colonial history, and he notes that it would be interesting to extend the comparison to the three other large Caribbean islands of Cuba, Jamaica, and Puerto Rico, at the cost of adding the complication of inter-island variation. Haber (Chapter 3) intentionally restricts his comparison of the development of banking systems after about A.D. 1800 to three New World countries (the United States, Brazil, and Mexico) and excludes European countries because all three of those New World countries began their independent existence without preexisting banks (their former colonial governments had not permitted the chartering of banks). Inclusion of European countries in the comparison would have introduced the complication of having to control for differences in bank development that already existed by 1800.

Another ubiquitous concern in natural experiments arises explicitly whenever one employs statistical tools for comparisons (though the concern is also implicit when one makes narrative comparisons without statistical tests). Does a statistical correlation by itself demonstrate a cause or a mechanism?

No, of course it doesn't: at least three further steps are required to demonstrate a cause or mechanism, and all three steps are the subjects of large methodological literatures. First, there is the problem of reverse causality: if A and B are correlated, perhaps A didn't cause B, as one assumed; perhaps, instead, B caused A. Frequently, one can approach this problem by examining time relations: in the simplest case, did A change before B, or vice versa? A statistical technique called Granger causality is often used to unravel the direction

of cause and effect. More sophisticated techniques are also employed. For instance, a recent study[3] identifies which brain regions stimulate which other brain regions when humans shift from relaxed to alert, and it does so by examining how phase differences between independent and dependent variables change with the frequency of their fluctuations.

Second, one must consider what is termed the *omitted variable bias:* the perturbing variable identified by the "experiment" may actually be part of a linked package of changes, within which some variable other than the one identified by the experimenter may really have been what caused the difference in outcomes. (This is essentially the concern that natural experimenters attempt to minimize, though without the possibility of complete success, as we described three paragraphs above.) Both Banerjee and Iyer in their study of the effects of the British colonial revenue system in India (Chapter 6), and Acemoglu et al. in their study of the effects of Napoleonic conquest (Chapter 7), wrestled with this problem. Among the many techniques that statisticians use to address this problem, an often-used technique is multiple regression analysis: that is, explicitly test the effects of other possible explanatory factors, and see whether the apparent explanatory power of one's initially preferred variable drops out when these other variables are taken into account.

Third, even if one has obtained convincing evidence that A causes B, further evidence is often required to establish the mechanism by which A causes B. For instance, human colonization of ecologically fragile Pacific islands is correlated with deforestation following human arrival, and it certainly is the case that human colonization somehow caused deforestation rather than that subsequent deforestation caused earlier human colonization. However, that observation by itself doesn't identify the mechanism by which human colonization resulted in deforestation. It could have involved direct actions by humans (such as people burning forests, chopping down trees, or using wood for fuel), or various indirect effects of humans (such as rats introduced by humans eating or gnawing on seeds of trees) Additional information that can help distinguish among these mechanisms includes archaeological and paleobotanical evidence of tree stumps with axe cuts, charcoal of identifiable tree species found in hearths, and nuts with gnaw marks left by rats' teeth.

In statistical analyses just as in narrative, noncomparative, nonquantitative historical studies, one has to negotiate a middle ground between overly simplistic and overly complex explanations. On the one hand, one might be concerned that statistical analysis would lead to oversimplified explanations, if one stopped looking for further explanatory factors after identifying the first couple of explanatory factors. In fact, statisticians attempt to add more independent variables to a multiple regression analysis, and they carry out residual analyses, in order to detect even more explanatory factors than emerged during the first stage of the analysis. Conversely, one may be suspicious about unnecessarily complex explanations, as expressed in the often-cited dismissive remark, "Give me two variables, and I will draw you an elephant; give me a third variable, and I will make him wave his trunk." In fact, statisticians routinely employ tests, such as the so-called F-test, in order to ascertain whether each additional variable tested really does add significant explanatory power beyond the power that one expects just from adding any randomly selected further variable.

In general, the more numerous are the potentially relevant independent variables, the more cases must be compared to test for effects of those variables. Conversely, the more cases that one has available for analysis, the greater the number of explanatory factors that can be tested. In this book the second largest scale comparison is Rolett's and Diamond's comparison of eighty-one Pacific islands or island sites in Chapter 4, examined for the outcome of deforestation. That large database made it possible to establish the existence of statistically significant and mechanistically understandable effects of nine independent variables: island rainfall, temperature, age, wind-borne

ash, wind-borne dust, makatea terrain, area, elevation, and isolation. Some of those effects were suggested to Rolett and Diamond by colleagues in the course of the study; the possible importance of these effects had not even occurred to Rolett and Diamond at the outset. With so many factors affecting deforestation, it would have been utterly impossible to evaluate them without a large database and without the use of statistics. Initially, Rolett and Diamond guessed—from their personal familiarity with two cases, the wet, warm, lightly deforested Marquesas Archipelago and the dry, cool, heavily deforested Easter Island—that rainfall and temperature would prove significant. While their full analysis did indeed confirm their hunch about the significance of rainfall and temperature, in retrospect that guess could not have been accepted based only on their initial narrative comparison of only two cases—because the Marquesas and Easter differ in other important respects as well.

But it is not true that a sufficiently large database will enable one to detect an effect of almost anything. For instance, Rolett and Diamond initially suspected that deforestation might also depend on variation in four agricultural practices: wet-field cultivation, dry-field cultivation, breadfruit arboriculture, and Tahitian chestnut and canarium arboriculture. But after expending two years of effort to tabulate the extent of each of these four practices on the eighty-one islands, Rolett and Diamond found no support for that initial hunch: none of these four agricultural practices had a statistically significant relationship to deforestation.

Social scientists have the misfortune of having to study fuzzier concepts than those studied by molecular biologists, physicists, chemists, and astronomers. The latter types of scholars aim to explain things that are easily defined, easily measured quantitatively, and often intuitively obvious—such as velocity, mass, chemical reaction rate, and luminosity. But we social scientists are interested in human happiness, motivation, success, stability, prosperity, and economic development. How does one build a meter to measure happiness? Human happiness is less neatly defined and harder to measure than molybdenum's atomic weight, but it is also more important to understand and explain.

Much of the practical difficulty in social science research resides in "operationalizing" fuzzy, hard-to-measure, but important concepts such as happiness. The scholar's task is to identify something that can be measured, and that can be shown to reflect or capture much of the essence of the ambiguous concept. For instance, historians interested in economic development today, at the touch of a computer button, can download vast, accurate databases of national incomes. But Acemoglu et al. (Chapter 7) want to understand whether Napoleon was good or bad for economic development in nineteenth-century Europe, at a time when incomes were not yet being measured and tabulated. What should they do? They resorted to "operationalizing" the fuzzy concept of economic development—that is, finding a proxy quantity which reflects economic development but a quantity for which data were already available in the early nineteenth century. A suitable proxy proves to be urbanization: specifically, the proportion of a region's population living in urban areas each containing 5,000 or more people. After searching for a proxy, economic historians have found this measure of urbanization useful because, historically, only regions with high agricultural productivity and a well-developed transport network—that is, areas fitting the fuzzy concept of "economically developed"—have been capable of supporting urban populations. Mathematicians and physical scientists who have never tried to measure something as important as urbanization or happiness often sneer at the efforts of social scientists to operationalize these concepts, and they quote examples of operationalizing pulled out of context in order to justify their scorn.[4]

What about the importance of quantitative data and measurements in historical studies?[5] In science in general, the role of quantification has been both overestimated and underestimated. As regards overestimation, quantification is so routinely essential in physics that

physicists have mistakenly assumed quantification to be essential to all of science. The great physicist Lord Kelvin wrote, "When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*." In fact, quantification played little role in the greatest advance in biology, Darwin's book *On the Origin of Species*. But while there are still some areas of sciences such as ethology and cultural anthropology in which one often begins by qualitative description, even in those areas it has become routine to go on to count a phenomenon's frequency or to describe it numerically. Insofar as possible, it helps to express in numbers the magnitude of effects and putative causes. Not only does that then permit numerical analyses, but it also forces a scholar to gather data more rigorously, and it furnishes objective measures that other scholars can check for themselves.

However, when scholars cannot express their effects and causes in numbers, they can still do many analyses merely by crudely ranking effect or cause magnitudes as weak, medium, or strong. For instance, although Rolett and Diamond (Chapter 4) were unable to put a number on Pacific island deforestation, they could still rank it on a qualitative five-point scale as negligible, mild, serious, very serious, or complete, and that enabled them to recognize the effects of nine influences or independent variables. Scholars in many other disciplines besides human history have to deal with nonnumerical variables, and many statistical tests developed to help those scholars will be useful to historians as well.

Whether one is able to express effects and causes in numbers or can only rank them crudely from weak to strong, one should try to assess apparent relations statistically. Such an assessment can not only help to protect one against the real risk that one's impressions about the main conclusions might prove to be wrong, but can also reveal other conclusions that one had not even suspected (as when

Rolett and Diamond were surprised to discover effects of island age, volcanic ash, and wind-borne Central Asian dust on Pacific island deforestation).

Every field of scholarship, not just human history, experiences tension between narrowly focused case studies and broader syntheses or generalizations. Practitioners of the case study method tend to decry syntheses as superficial, coarse-grained, and absurdly oversimplified; practitioners of syntheses tend to decry the case studies as merely descriptive, devoid of explanatory power, and unable to illuminate anything except one particular case study. Eventually, scholars in mature fields come to realize that scholarly understanding requires both approaches. Without reliable case studies, generalists have nothing to synthesize; without sound syntheses, specialists lack a framework within which to place their case studies. Thus, comparative history poses no threat to the more familiar approach of historical case studies, but on the contrary offers a means to enrich that approach.

The tension between case studies and syntheses, or between description and theoretical explanation, has unfolded differently in different fields of scholarship. This tension is minimal in physics and chemistry, where theoreticians and experimentalists now take it for granted that each needs the other, and where it is now routine to place narrow case studies within a larger framework. Among scholarly fields that use natural experiments rather than manipulative experiments, there has been recent tension between the two approaches especially in cultural anthropology and field biology. Cultural anthropologists used to view each human culture as unique and therefore resisted generalization. But today virtually every anthropologist publishing the results of a multiyear study of some particular tribe will begin the publication with a section developing some general theoretical perspective and placing that tribe along a spectrum of cultural variation.

In the field of ecology, tension between case studies and generalization became acute in the 1960s and 1970s, with the development of

many new theoretical generalizations and mathematical models. That development gave rise to nearly twenty years of bitter disputes. On the one side were the traditional field biologists who had devoted their lives to long-term studies of one animal or plant species, such as the Philippine Striped Tit-Babbler. Attempts to compare, model, theorize, and generalize were derided with labels such as "superficial," "oversimplified," and "generalizations based on caricatures without the rich detail of my study of Philippine Striped Tit-Babblers." These scholars warned other scientists that progress could come only through equally richly textured, carefully nuanced studies of other bird species. On the other side, theorizing generalists began to object, "You can't hope to understand even just the Philippine Striped Tit-Babbler, without understanding how and why it became similar to and different from other tit-babblers and other bird species."

Within ecology, today, the polar approaches of case studies and generalization coexist more comfortably.[6] Most ecologists now recognize that their discipline is developing a general framework that applies to species as diverse as bacteria, dandelions, and woodpeckers—a framework that allows an understanding of differences within the plant and animal kingdoms. It is no longer enough to describe how one bird does this, while another bird does that. One after another, the leading bird journals, although still publishing accounts of individual bird species, have come to *require* that each study be placed within a larger framework.

Setting individual explanations within a larger explanatory framework is a hallmark of science. For example, Darwin noticed that the mockingbirds of the Galapagos Islands were related to South American mockingbirds, but he also noticed that other Galapagos species as well have their closest relatives in South America. Such observations stimulated Darwin and Wallace to set those facts into a larger framework of biogeographic explanation, which combined history, dispersal, evolution, and origins or movements of land masses. Chemists studying the molybdenum atom don't explain it as a unique phenomenon but fit its properties into an explanatory framework based on the periodic table, atomic theory, and quantum mechanics.

The case studies of this book support two overall conclusions about the study of human history. First, historical comparisons, though not providing all the answers by themselves, may yield insights that cannot be extracted from a single case study alone. For instance, one cannot hope to understand late nineteenth-century France without examining why it differed from late nineteenth-century Germany or late sixteenth-century France. Second, insofar as is possible, when one proposes a conclusion, one may be able to strengthen that conclusion by gathering quantitative evidence (or at least ranking one's outcomes from big to small), and then by testing the conclusion's validity statistically.

Some specialist historians would respond with an implicit objection, which is sometimes but not always expressed openly, and which we mentioned in the prologue. An example of this objection could be phrased as follows: "I have devoted forty years of my professional life to studying the American Civil War, and I still don't fully understand it. How could I dare to discuss civil wars in general, or even just to compare the American Civil War with the Spanish Civil War, to which I have not devoted forty years of study? And, worse yet, isn't it outrageous that some scholar of the Spanish Civil War dares to trespass on my turf and to say something about the American Civil War?" Yes, if you study an event for a long time, that does give you one type of advantage. But you gain a different type of advantage by taking a fresh look at an event, and by applying to it the experience and insights that you have gained by studying other events. We hope that this book will offer useful guidelines to historians and social scientists desiring to exploit that advantage.

NOTES

1. The books and papers cited in note 3 of the prologue will be useful for further discussion of the problems discussed in this afterword.

2. David Landes, *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present* (Cambridge, 1969); Douglass North and Robert Thomas, *The Rise of the Western World: A New Economic History* (New York, 1973); E. L. Jones, *The European Miracle: Environments, Economies, and Geopolitics in the History of Europe and Asia*, 2nd ed. (Cambridge, 1987); Graeme Lang, "State Systems and the Origins of Modern Science: A Comparison of Europe and China," *East-West Dialog* 2 (1997): 16–30; Kenneth Pomeranz, *The Great Divergence: China, Europe, and the Making of the Modern World Economy* (Princeton, NJ, 2000); Angus Maddison, *The World Economy: A Millenial Perspective* (Paris, 2001); Jack Goldstone, "Efflorescences and Economic Growth and World History: Rethinking the 'Rise of the West' and the Industrial Revolution," *Journal of World History* 13 (2002): 329–389; Joel Mokyr, *The Enlightened Economy: An Economic History of Britain, 1700–1850* (New Haven, CT, 2007); Jan Luiten van Zanden, "Die mittelalterlichen Ursprünge des 'europäischen Wunders,'" in James Robinson and Klaus Wiegandt, eds., *Die Ursprünge der Modernen Welt* (Frankfurt am Main, 2008), pp. 475–515; Michael Mitterauer, "Mittelalterliche Wurzeln des europäischen Entwicklungsvorsprungs," in James Robinson and Klaus Wiegandt, eds., *Die Ursprünge der Modernen Welt* (Frankfurt am Main, 2008), pp. 516–538.
3. G. Nolte et al., "Robustly Estimating the Flow Direction of Information in Complex Physical Systems," *Physical Review Letters* 100 (2008): 234101-1–234101-4.
4. Jared Diamond, "Soft Sciences Are Often Harder than Hard Sciences," *Discover* 8, no. 8 (1987): 34–39.
5. Some of this discussion is drawn from a chapter by Jared Diamond, "Die Naturwissenschaft, die Geschichte und Rotbrustige Saftsäuger," in Robinson and Wiegandt, eds., *Die Ursprünge der Modernen Welt*, pp. 45–70.
6. Robert May and Angela McLean, *Theoretical Ecology*, 3rd ed. (Oxford, 2007).