Natural Experiments of History

EDITED BY

Jared Diamond

James A. Robinson

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

Prologue

JARED DIAMOND AND
JAMES A. ROBINSON

The controlled and replicated laboratory experiment, in which the experimenter directly manipulates variables, is often considered the hallmark of the scientific method. It is virtually the only method employed in laboratory physical sciences and in molecular biology. Without question, this approach is uniquely powerful in establishing chains of cause and effect. That fact misleads laboratory scientists into looking down on fields of science that cannot employ manipulative experiments.

But the cruel reality is that manipulative experiments are impossible in many fields widely admitted to be sciences. That impossibility holds for any science concerned with the past, such as evolutionary biology, paleontology, epidemiology, historical geology, and astronomy; one cannot manipulate the past.[1] In addition, when one is studying bird communities, dinosaurs, smallpox epidemics, glaciers, or other planets, manipulative experiments that are possible in the present would often be condemned as immoral and illegal; one should not kill birds or melt glaciers. One therefore has to devise other methods of "doing science": that is, of observing, describing, and explaining the real world, and of setting the individual explanations within a larger framework.

A technique that frequently proves fruitful in these historical disciplines is the so-called natural experiment or the comparative

method. This approach consists of comparing—preferably quantitatively and aided by statistical analyses—different systems that are similar in many respects but that differ with respect to the factors whose influence one wishes to study. For instance, to study the ecological effect of woodpeckers known as Red-breasted Sapsuckers on related woodpeckers known as Williamson's Sapsuckers, one can compare mountains, all of which support Williamson's Sapsuckers but some of which support Red-breasted Sapsuckers while others do not. The science of epidemiology is virtually the study of such natural experiments on human populations. As one example, we have learned which human blood groups provide resistance to smallpox, not as a result of manipulative experiments in which we inject people carrying different blood groups either with smallpox virus or with a virus-free control solution, but instead as a result of observations of people carrying different blood groups during one of the last natural smallpox epidemics in India several decades ago. Physicians who were present in a remote village at the time of the outbreak determined villagers' blood groups and observed who got sick or died and who did not.[2]

Of course, natural experiments involve many obvious pitfalls. These pitfalls include the risk that the outcome might depend on other factors that the "experimenter" had not thought to measure; and the risk that the true explanatory factors might be ones merely correlated with the measured factors, rather than being the measured factors themselves. These and other such difficulties are real—but so are the difficulties encountered in executing manipulative laboratory experiments or in writing noncomparative narrative accounts. An extensive literature is now available on how best to overcome these pitfalls.[3]

For example, consider a question that is currently of much practical interest: does smoking cause cancer? It is possible to write a moving, nuanced, in-depth biography of one particular smoker who did die of cancer, but that narrative doesn't prove that smoking causes cancer in general or even that it caused that particular cancer. Some smokers don't get cancer, and some nonsmokers do get it. As we have learned, there are many other risk factors for cancer besides smoking.

Hence epidemiologists routinely gather data on thousands or millions of individuals, code them not only for whether they smoke but also for their diet and many other factors, and then carry out a statistical analysis. Such studies yield familiar and now widely accepted conclusions. Yes, smoking is strongly associated with some (though not with other) forms of cancer, but one can also recognize many other causes by means of statistical analyses. Those other causes include dietary fat, dietary fiber, dietary antioxidants, sun exposure, individual air pollutants, specific chemicals in our food and water, numerous hormones, and hundreds of different genes. Hence no epidemiologist would dream of identifying *the* cause of cancer just by telling the story of a single patient, but one can convincingly identify many causes of cancer by comparing and statistically analyzing many people. Similar conclusions and similar pitfalls that need to be addressed apply to multicausal historical phenomena.

On reflection, one might also expect comparisons and quantitative methods and statistics to play an uncontroversial middle role in the study of history. Historians are constantly making statements of the form "This changed (or increased or decreased) with time," or "This was more than that," or "This person did more (or less) than, or behaved differently from, that person." But merely to make such statements, without providing the underlying numbers and doing the associated statistics, is to frame the comparison without carrying it out. Already in 1979, the historian Lawrence Stone made this same point in his discussion of the role of quantification: "Historians can no longer get away with saying 'more', 'less', 'growing', 'declining', all of which logically imply numerical comparisons, without ever stating explicitly the statistical basis for their assertions. It [quantification] has also made argument exclusively by example seem somewhat disreputable. Critics now demand supporting statistical evidence to show that the examples are typical, and not exceptions to the rule."[4]

In reality, the various social sciences concerned with human societies have made uneven use of natural experiments. Although

there is widespread acceptance of natural experiments in archaeology, cultural anthropology, developmental psychology, economics, economic history, political science, and sociology, in the field of human history other than economic history their use has been patchy. Some historians merely call for more use of natural experiments; others claim that other historians already do use them a lot; and still others actually do use them, though sometimes not consciously or without making full use of the methodological advantages potentially associated with this approach.[5] But many historians do not use natural experiments at all and are skeptical or hostile to the approach, especially to systematic comparisons involving quantitative data that are analyzed statistically.

Numerous reasons contribute to this skepticism. One reason is that the discipline of history is variously grouped either with the humanities or with the sciences. At one major American university, for instance, the undergraduate college places the history department under the dean of humanities, but the graduate school places it under the dean of social sciences. Many students who choose to train as historians rather than as economists and political scientists do so explicitly to avoid having to learn mathematics and statistics. Historians often devote their careers to studying one country or geographic region within one slice of time. The special expertise required to master that region and period leads its students to doubt that a historian who has not spent his or her life acquiring that expertise could write knowledgeably about that region and period, or that they themselves could knowledgeably compare it with a different region and period. The lengthy training required of graduate students in history involves strong socialization about what history is and is not, and about what methods are or are not proper for historians. Many American historians reacted to the debate initiated by a particular school of quantitative history, termed cliometrics, by becoming less quantitative—as if the weaknesses claimed by critics of this particular approach applied to all quantitative analyses.[6] Historians often believe that human history is fundamentally different

from the history of cancers, chimpanzees, or glaciers, on the grounds that it is much more complicated and involves the motives of individual humans, which supposedly cannot be measured or expressed in numbers. However, cancers, chimpanzees, and glaciers are also very complicated, and they pose the added obstacle that they do not leave behind any written archival evidence of their motives. In addition, many scholars, such as psychologists, economists, scholars of government, and some biographers, now are able to measure and analyze the motives of individual humans by means of retrospective analyses of documents of dead people as well as interviews with still-living people.

Our book seeks to showcase the comparative method in history and to examine some techniques for solving its obvious pitfalls by presenting a set of eight studies in seven chapters (Chapter 4 includes two studies). Our target audience is not just those historians receptive to (or at least not implacably opposed to) the comparative method, but also the larger number of scholars in allied social sciences that already widely employ the comparative method. We write for undergraduates as well as for established scholars. We do not assume familiarity with statistics or quantitative analyses. The eight studies (two of them coauthored) are by eleven authors, two of whom are traditional historians based in history departments, while the others are drawn from archaeology, business studies, economics, economic history, geography, and political science. These studies are designed to cover a spectrum of approaches to comparative history, in four respects:

First, the approaches range from a nonquantitative narrative style traditional among historians, in the early chapters, to quantitative studies with statistical analyses familiar in the social sciences outside history departments, in the later chapters.

Second, our comparisons range from a simple two-way comparison (the nations of Haiti and the Dominican Republic sharing the island of Hispaniola) to three-way comparisons in two chapters,

through comparisons of dozens of German regions, up to comparisons of 81 Pacific islands and 233 areas of India.

Third, the societies that we study range from contemporary ones, through literate societies of recent centuries for which we have abundant written archival information, to nonliterate past societies for which all our information comes from archaeological excavations.

Finally, our geographic coverage offers something for historians of many different parts of the world. Our case studies encompass the United States, Mexico, a Caribbean island, Brazil, Argentina, Western Europe, tropical Africa, India, Siberia, Australia, New Zealand, and other Pacific islands.

Traditional historians will thus find the approach of the first four studies in this book familiar in that they develop evidence in a narrative style, compare small numbers of societies (three, seven, three, and two, respectively), and do not present statistical comparisons of quantitative data in the text. The approach of the remaining four studies differs from that of most traditional historians but will be familiar to some historians and to scholars in related social sciences, in that they are explicitly based on statistical comparisons of quantitative data and they compare many societies (81, 52, 233, and 29, respectively).

In Chapter 1, Patrick Kirch asks why history unfolded so differently among the dozens of Pacific islands colonized by a single ancestral people, the early Polynesians. Kirch focuses on three islands or archipelagoes spanning the range of sociopolitical and economic complexity in Polynesia: the small island of Mangaia, which developed as a small-scale chiefdom; the medium-sized Marquesas archipelago, which came to support multiple independent warring chiefdoms; and Hawai'i, the largest Polynesian archipelago outside New Zealand, which developed several large-scale competing polities characterized as emerging "archaic states," with each occupying one or more islands. Because all of those Polynesian societies lacked writing, Kirch's study rests on linguistic, archaeological, and ethnographic evidence rather than on the written archival evidence emphasized

by historians. Kirch's research is therefore conventionally labeled as archaeology rather than as history, although his questions are ones familiar to traditional historians. Kirch notes that similarities in cultural traits among societies may arise either through parallel retention of the same ancestral trait (so-called shared homologies), independent development (so-called analogies), or borrowing. Hence Kirch sets out a methodologically rigorous approach to comparisons that he terms the phylogenetic model, and he uses multiple lines of evidence (the "triangulation" approach) to reconstruct aspects of past societies and cultures.

James Belich (Chapter 2) adds to the extensive literature on frontier societies, such as those of the American West, by comparing seven such nineteenth-century societies: those in the United States, the "British Wests" (Canada, Australia, New Zealand, and South Africa), Argentina, and Siberia. These societies differed in many obvious respects, such as in their proportion of immigrants who returned to the mother country; in their decade of maximum growth and hence the prevailing stage of the Industrial Revolution; and especially in that five of the societies were Anglophone, one (Argentina) was Spanish-speaking but received even more Italian than Spanish immigrants, and one (Siberia) was Russian. Despite those different "experimental conditions," Belich's most striking conclusion is that all of those frontiers repeatedly traversed similar three-step cycles of an explosive population boom marked by net imports of goods and capital, then a dramatic "bust" decimating growth rates and bankrupting farms and businesses, and finally an export rescue creating a new economy based on mass export of staples to a distant metropolis. Belich documents a total of twenty-six such cycles on his seven frontiers. Their repeated emergence suggests that the underlying similarities of population and economic dynamics of all of those frontiers overrode the influences of their differences in immigrant commitment, decade of growth, stage of industrialization, and mother country. More generally, Belich's results illustrate that students of comparisons must be alert not only to differences but also to similarities in

outcomes: *convergent evolution*, to borrow a term from evolutionary biology.

Stephen Haber (Chapter 3) compares the United States, Mexico, and Brazil with respect to the nineteenth-century origins of their banking systems, whose differences had heavy consequences for the subsequent modern histories of those three countries. Haber's case study contributes to a general question that has been much studied by economists, political scientists, and historians: why do some countries have large banking systems that allocate credit broadly, thereby permitting rapid growth, while other countries have scarcely any banks at all, thereby constraining growth and limiting societal mobility? As an example of national differences, in the year 2005 private bank loans equaled 155% of gross domestic product in the United Kingdom, 98% in Japan, 15% in Mexico, and 4% in Sierra Leone. Those national differences in banking systems are obviously related to differences in democratic governance, but that raises the question of the direction of causation: do democratic institutions promote large banking systems, or, conversely, do large banking systems promote democratic institutions? To reduce confounding variables in his natural experiment, Haber selects three large New World countries, all of which obtained their independence within a few decades before or after 1800, and all of which started nationhood with no chartered banks (because their former European colonial rulers had forbidden them). That selection by Haber reduces the complications that would have been encountered by extending the study to European countries, which already had chartered banks (and important differences in their banking systems) as of 1800. Each of the three New World countries chosen provides smaller internal natural experiments embedded within a larger natural experiment: not only did they differ in their political institutions, but also those institutions in each country changed over time during the era studied (from independence until roughly 1914).

In the last and smallest-scale study among our four narrative nonstatistical case studies, Jared Diamond (Chapter 4) compares two

societies—Haiti and the Dominican Republic—that divide the Caribbean island of Hispaniola across one of the most dramatic political boundaries in the world. Viewed from an airplane, Hispaniola is bisected by a sharp line: to the west, the brown, treeless expanse of Haiti, heavily eroded and more than 99% deforested; to the east, the green of the Dominican Republic, still nearly one-third covered with forests. The political and economic differences between these two countries are equally stark: densely populated Haiti is the poorest country in the New World, with a weak government unable to provide basic services to most of its citizens, while the Dominican Republic, though still a developing country, has an average per capita income six times that of Haiti, many export industries, and a recent succession of democratically elected governments. A small part of those differences between modern Haiti and the Dominican Republic is due to differing initial environmental conditions: Haiti is somewhat drier and steeper, and has thinner and less fertile soils, than the Dominican Republic. But the largest part of the explanation lies in their colonial histories: western Hispaniola became colonized by France, eastern Hispaniola by Spain. That difference in colonial power initially produced major differences in slave plantations, language, population density, social inequality, colonial wealth, and deforestation, leading first to differences in the struggle for independence; then to differences in receptivity to foreign investment and immigration, and differences in perception by Europeans and Americans; more recently, to differing modern long-lasting dictators; and finally to the different conditions of these two countries today.

The other study of Chapter 4 goes to the opposite extreme: after that small-scale narrative comparison of the two halves of a single island, we consider a large-scale statistical comparison of sixty-nine Pacific islands, and of the wet and dry parts of twelve of those islands. The starting point of this study is the romantic mystery of Easter Island, famous for its hundreds of toppled giant stone statues: why did Easter Island end up as the Pacific's most deforested island, with virtually all of its native tree species extinct, and with

heavy consequences for its wood-dependent human society? But Easter Island is just one data point in a larger natural experiment, since deforestation among the Pacific's hundreds of islands ranged from complete (as on Easter) to negligible. Diamond's database includes the islands studied by Kirch in Chapter 1 and settled by Polynesians, as well as islands settled by two related groupings of Pacific peoples (Melanesians and Micronesians). Because tree growth and deforestation depend on many factors, it would have been impossible for a narrative study of just one or two islands to help us understand this range of outcomes. But the large number of islands available for analysis makes it possible to identify significant influences on deforestation from nine separate factors, several of which Diamond and his collaborator Barry Rolett did not even imagine might be important until they carried out their statistical analyses. Of wider interest to historians was the possibility of extracting these conclusions even without measuring deforestation quantitatively: Rolett and Diamond only ranked it crudely on a five-point scale from severe to mild. Historians often seek to understand outcomes that are difficult to measure but that can at least be ranked ("big," "medium," "small"). Those historians can make use of the whole branch of statistics devoted to analyzing such ranked nonnumerical outcomes.

The remaining three studies—by Nathan Nunn (Chapter 5), Abhijit Banerjee and Lakshmi Iyer (Chapter 6), and Daron Acemoglu, Davide Cantoni, Simon Johnson, and James Robinson (Chapter 7)—all describe natural experiments in which the historical consequences of some massive perturbation (respectively, the African slave trade, British colonial rule in India, and institutional changes accompanying French Revolutionary conquests) can be examined because the perturbation operated in a geographically irregular patchwork over a large region. When one compares the perturbed patches with the unperturbed patches, it is thus a plausible hypothesis, worth testing, that average societal differences observed between the two types of patches arose from the operation or nonoperation of the perturbing

factor rather than from some other differences between the patches. If, however, the patches with and without the factor had instead been distributed in some geographically regular way (e.g., all the patches with the factor being in the south or at high altitude), it would have been an equally plausible hypothesis that those geographic differences rather than the presence or absence of the factor caused the observed societal differences. Of course, all three studies must also address the question of the direction of cause and effect: did the perturbations really cause the observed differences, or might the instigators of the perturbations (respectively, the slave traders, British administrators, and French conquerors) have instead chosen particular patches in a geographically irregular patchwork because of preexisting differences that should be considered the real causes of the modern differences?

One of those three studies, Nathan Nunn's, explores the long-standing question of the slave trade's legacies for modern Africa, by comparing modern African states whose territories experienced differing impacts in the past from the slave trades across the Atlantic Ocean, the Sahara, the Red Sea, and the Indian Ocean. Many slaves were exported from some parts of Africa, while virtually no slaves were taken from other parts. Today, the former slave-exporting parts tend to be poorer than the former non–slave-exporting parts, and Nunn argues that the slave trades caused the economic differences rather than vice versa. Similarly, Abhijit Banerjee and Lakshmi Iyer address the unresolved question about the impact of British colonial rule on India. They find that areas of India directly administered by the British colonial government in the past tend today to have fewer schools and paved roads, lower literacy, and less use of domestic electricity than areas indirectly administered in the past. Similarly again, Daron Acemoglu, Davide Cantoni, Simon Johnson, and James Robinson explore the debate concerning the effects of the massive institutional changes introduced by French Revolutionary armies and Napoleon into conquered areas of Europe. The authors compare areas of Germany with and without the massive institutional changes,

and they describe the historical accidents that caused the changes to be applied in a geographically irregular patchwork over Germany. Those institutional changes led to increased urbanization—but only after a lag of several decades, owing to the lag in arrival of the Industrial Revolution. Whereas areas that had experienced institutional changes embraced the Industrial Revolution, areas that had clung to their old institutions resisted it.

A concluding afterword reflects on methodological issues common to these and other studies of natural experiments of human history by comparative methods. Those issues include natural experiments involving either different perturbations or different initial conditions; the "selection" of sites that were perturbed; time lags for effects of perturbations to emerge; problems in inferring causality from an observed statistical correlation, such as questions of reversed causality, omitted variable bias, and underlying mechanisms; methods for steering between the opposite traps of overly simplistic and overly complex explanations; "operationalizing" fuzzy phenomena (e.g., how to measure and study happiness); the role of quantification and statistics; and the tension between narrow case studies and broader syntheses.

With regard to our book's style and format, we recognize that most multiauthored volumes suffer from having too many chapters and authors, too many pages, too little unity, and too little editing. Both of us have edited at least two multiauthored volumes, and we know painfully well the effort required to achieve a well-integrated result. We calculate that our urgings of the coauthors of those completed volumes cost us on the average, per volume, two friendships for life and several more friendships for at least a decade. Fortunately, all of our current authors have read all of each other's drafts, and in the present case all have remained gracefully cooperative in responding to our endless requests for revision over the two years that we have been working on this project. Each chapter has also been read by a half dozen traditional historians, whose suggestions we have incorporated or taken into account.[7]

NOTES

It is a pleasure to acknowledge our debts to Robert Schneider and his colleagues, to many others of our own colleagues, and to many anonymous as well as signed reviewers, for their generosity with their time and for their suggestions, which helped shape and improve this book.

1. Ernst Mayr has written thoughtfully about differences between historical and nonhistorical sciences. See, for instance, Ernst Mayr, *This Is Biology: The Science of the Living World* (Cambridge, MA, 1997).
2. F. Vogel and N. Chakravartti, "ABO Blood Groups and Smallpox in a Rural Population of West Bengal and Bihar (India)," *Human Genetics* 3 (1966): 166–180.
3. Discussions of the pitfalls in inferring causes from natural experiments include Jared Diamond, "Overview: Laboratory Experiments, Field Experiments, and Natural Experiments," in Jared Diamond and Ted Case, eds., *Community Ecology* (New York, 1986), pp. 3–22; William Shadish, Thomas Cook, and Donald Campbell, *Experimental and Quasi-experimental Designs for Generalized Causal Inference* (Boston, 2002); James Mahoney and Dietrich Rueschemeyer, eds., *Comparative Historical Analysis in the Social Sciences* (New York, 2003); Joshua Angrist and Jorn-Steffan Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton, NJ, 2008); Guido Imbens and Donald Rubin, *Causal Inference in Statistics, and in the Social and Biomedical Sciences* (Cambridge, 2008); and Thad Dunning, "Improving Causal Inference: Strengths and Limitations of Natural Experiments," *Political Research Quarterly* 61 (2008): 282–293.
4. Lawrence Stone, "The Revival of Narrative: Reflections on a New Old History," *Past and Present*, no. 85 (1979): 3–24, quotation pp. 10–11.
5. An example might be the debate initiated by Robert Brenner's paper "Agrarian Class Structure and Economic Development in Preindustrial Europe," *Past and Present*, no. 70 (1976): 30–75. Papers in the debate were collected by T. H. Aston and C. H. E. Philpin, eds., *Agrarian Class Structure and Economic Development in Pre-industrial Europe* (New York, 1987). The debate concerned why the Black Death had such different consequences in western and eastern Europe. To use the terminology that we shall explain in the afterword of this volume, the debate examined how a common perturbation led to different consequences in different areas as a result of different initial conditions.

6. The debate over cliometrics was explored by Robert William Fogel and G. R. Elton, *Which Road to the Past? Two Views of History* (New Haven, CT, 1983).
7. Some of this discussion is drawn from a chapter by Jared Diamond, "Die Naturwissenschaft, die Geschichte und Rotbrustige Saftsäuger," in James Robinson and Klaus Wiegandt, eds., *Die Ursprünge der Modernen Welt* (Frankfurt am Main, 2008), pp. 45–70.

Controlled Comparison and Polynesian Cultural Evolution

PATRICK V. KIRCH

In early January of 1778, Captain James Cook, in command of HMS *Resolution* and *Discovery*, was sailing through uncharted waters in the central North Pacific Ocean, en route to the coast of New Albion, as the Pacific Northwest was then called. The Admiralty had instructed Cook to replenish at Tahiti, an island he already knew well from two previous voyages, then to go northward in search of the fabled "northwest passage." On January 18, the *Resolution*'s lookout spied a high island to the northeast; a second volcanic peak was soon discerned to the north. The following day Cook and his crew made "first contact" with one of the most isolated societies on earth—the Polynesian inhabitants of Kaua'i, one of the Hawaiian Islands.

Cook was no stranger to Polynesia. He had first gone to Tahiti a decade earlier, at the behest of the Royal Society of London, to observe the June 3, 1769, transit of Venus across the sun. That mission accomplished, Cook extended his explorations to other islands of the Society archipelago, followed by an unprecedented circumnavigation of New Zealand. In 1772 the Admiralty dispatched him again to the Pacific, to determine whether or not the long-hypothesized continent of Terra Australis actually existed. In addition to taking his ships farther south than any man had gone before, Cook explored and mapped more of Polynesia, including the Tuamotu Islands, Tonga, the southern Cook Islands, Easter Island, and the Marquesas.