

Review of *Natural Experiments of History*

Thad Dunning

Department of Political Science

Yale University

[Prepared for publication in *Perspectives on Politics*]

This draft: June 1, 2010

Diamond, Jared, and James Robinson, eds. 2010. *Natural Experiments of History*. Cambridge, MA: The Belknap Press of Harvard University Press.

The contributions to *Natural Experiments of History*, a striking new volume edited by Jared Diamond and James A. Robinson, address substantive topics as diverse as the economic impact of the African slave trade, the divergent evolution of polities in Polynesia, and the effects of colonial land tenure systems in India. Drawing together work from anthropology, economics, geography, history, and political science, these diverse essays nonetheless undertake a similarly difficult task: that of drawing causal inferences from the historical record.

They do so compellingly, and in very interesting fashion. Thus, Nathan Nunn's chapter shows that current-day African countries that exported more slaves per capita, in the trans-Atlantic as well as the smaller trans-Saharan, Red Sea, and Indian Ocean slave trades, are poorer today than African countries where little or no slave trade existed. This pattern moreover reflects a striking reversal of fortune: using data on population density in 1400 as a proxy for economic development,

Nunn shows that slave-exporting areas were in fact richer prior to the start of the slave trade than non-exporters. Nunn draws on data that are the fruit of painstaking efforts by many historians of the slave trade, using export records in African ports as well as bills of sale, plantation inventories, slave registers, and many other New World sources to estimate the numbers and ethnic identities (and thus the approximate geographic origins) of African slaves. Nunn interprets his findings as evidence for the corrosive impact of the slave trade on formal and informal institutions, though evidence on this point is here somewhat thin. Issues of interpretation aside, the essay presents striking documentation of the broad historical pattern—in an area in which systematic comparative research is especially difficult.

Other contributions to the volume reach similarly interesting conclusions. Abhijit Banerjee and Lakshmi Iyer argue that areas of colonial India subject to landlord-based revenue systems—in which the British collected taxes indirectly through powerful landlords, rather than directly from individual cultivators or village bodies—tended to have lower levels of investment in primary education, electricity, and roads as late as 1991, long after these revenue systems had been abolished in independent India. They interpret this finding as evidence for the corrosive developmental impact of powerful vested interests, though their data on patterns of electoral competition does not strongly support the idea of “political capture” by landlords. Daron Acemoglu, Davide Cantoni, Simon Johnson, and James A. Robinson argue that areas of Germany invaded by Napoleon after the French Revolution—and thus often subjected to a range of institutional innovations, from

the civil code to the abolition of guilds—prospered economically, particularly after 1850, relative to non-invaded areas; the best economic performers were areas claimed by Prussia in the Congress of Vienna, where Napoleon’s main institutional innovations were maintained. Stephen Haber examines the divergence of capital markets in post-colonial Brazil, Mexico, and the United States, attributing the eventual development of a broader banking system in the U.S. to that country’s relatively democratic dispersal of political power, as well as to competition among the states and between the states and the federal government.

The remaining contributions to the volume also emphasize the value of historical comparison. Thus, Patrick V. Kirch attributes the diverse political trajectories of three island regions in Eastern Polynesia—Mangaia, the Marquesas, and Hawai’i, all settled by allegedly similar groups of émigrés from Western Polynesia late in the first millennium A.D.—to the different geological and agricultural conditions encountered by the settlers in each region. James Belich finds a common three-stage cycle of boom and bust followed by an “export rescue” in the frontier zones of such diverse countries as the United States, Canada, Australia, New Zealand, South Africa, and Argentina. Finally, Jared Diamond examines the disparate developmental trajectories of the Dominican Republic and Haiti—two nations sharing the island of Hispaniola but separated by a border—and separately compares the collapse of Polynesian society on Easter Island to the persistence of Polynesian societies elsewhere in the Pacific.

The broader methodological agenda of the book is thus to showcase the use of the comparative method in history. For purposes of drawing causal inferences, of course, comparisons are most useful insofar as the group of units exposed to a putative cause—the African slave trade, the colonial land revenue system, and so on—are as similar as possible to the group of units not exposed to the cause, in all respects except for the presence or absence of the cause. Indeed, comparison may be most powerful when the putative causal factors are as good as randomly assigned, and thus confounding variables that might otherwise explain ex-post differences across groups do not vary systematically with exposure to the cause—as in the so-called “natural experiments” that give this volume its title.

A difficulty with natural experiments is that they are generally challenging to validate. History does not often divide units into “treatment” and “control” groups through a true randomizing device such as a lottery. Instead, as Diamond and Robinson put it in their methodological conclusion to the volume, “important historical decisions are rarely made by flipping coins: Napoleon [had] his reasons for invading certain German principalities but not others...just as slave traders had their own reasons for buying slaves from certain parts of Africa but not others. 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 respect to that outcome)?” (Diamond and Robinson 2010: p. 262). Achieving the ex-ante comparability afforded by random assignment is seriously challenging in historical studies, since history rarely provides the degree of control afforded by a laboratory environment or a field experiment.

Yet if natural experiments are difficult to validate—since the assertion of *as-if* random assignment is ultimately only partially testable—it is certainly possible to *invalidate* an alleged natural experiment. As a necessary if not sufficient condition, in a valid natural experiment, the treatment and control groups should be balanced with respect to “pre-treatment covariates” or what Diamond and Robinson call “initial conditions”—just as they would be if the treatment were truly assigned at random. A variety of quantitative and qualitative evidence should also be mustered to support the assertion that the process governing selection into treatment or control groups is akin to a coin flip. While provision of such evidence does not guarantee the existence of a valid natural experiment, evidence to the contrary—such as substantial differences in initial conditions—should raise warning signs.

In many of the studies featured in the volume, it is difficult to argue that treatment assignment is as good as random. In Nunn’s very compelling study of the African slave trade, for example, the fact that slaves came disproportionately from societies that were richer in 1400 might contribute to our confidence that the slave trade sparked a reversal of fortune—but it also suggests that the treatment was hardly allocated *as-if* at random. Indeed, as Nunn points out, European slave traders sought out densely-populated and wealthy societies, as it was easier to capture and trade slaves from such societies. One difficulty here is that local conditions that were propitious for African economic development in the historical period before the slave trade might be distinct from those that promote development today—and we cannot easily distinguish the effects of such disparate initial conditions from the effects of the slave trade, since they co-vary. There are also non-trivial threats to the

claim of *as-if* random assignment in other studies; for example, the relatively democratic dispersal of political power in the United States may well be related to other factors that explain broad-based banking systems. Napoleon may have invaded Rhineland and other parts of Western Germany for reasons unrelated to their economic potential, but these areas also appear to differ systematically from the landlord-dominated areas in the East of the country. Finally, in several other studies in the volume (e.g., Belich), a shared pattern is identified across units with some commonalities, but such studies do not appear to constitute natural experiments—that is, settings in which a treatment or putative cause has been applied *as-if* at random to some units but not others.

The possible absence of *as-if* random assignment in no way invalidates the conclusions of any particular study, of course; as Diamond and Robinson emphasize in their methodological Afterword, it simply implies that researchers must worry all the more about common threats to valid causal inference in observational (non-experimental) settings. This careful concern is on impressive display in many of these essays. In companion pieces to those published in this volume, several of the authors attempt to control for potential confounders—for example, through the introduction of covariates in a multivariate regression analysis, or the use of instrumental-variables analysis—and their main conclusions appear convincing. Yet, it should also be emphasized that controlling for confounders through statistical adjustment is quite different than doing so through research design. Indeed, the use of statistical fixes like multiple regression analysis should perhaps be viewed as an admission that these are not really natural experiments of history—after all, if the

presumed cause is truly assigned *as-if* at random, controlling for potential confounders is typically unnecessary.<sup>1</sup>

What this landmark volume really shows, convincingly and stunningly, is the great value of sustained historical comparison, conducted by researchers who leverage various kinds of evidence and pay careful attention to issues of causal inference. Drawing causal lessons from the historical record is certainly challenging. Yet, this volume compellingly demonstrates that carefully chosen comparisons—whether qualitative or quantitative—can increase our confidence in causal inferences across a very diverse set of substantive topics and sociopolitical settings. The contributors all discuss substantive areas in which it is relatively challenging—and thus all the more valuable—to draw conclusions about social and political causes and effects. More researchers should therefore pay heed to the broad, interdisciplinary, and ambitious agenda outlined by this impressive book.

---

<sup>1</sup> Researchers sometimes estimate multiple regression models on (natural) experimental data to reduce the variance of treatment effect estimators, though variance can in general be larger or smaller after adjustment.