

# Economics and Natural Experiments

---

*Dossier*



# Natural Experiments and Causality in Economic History

## On Their Relations to Theory and Temporality\*

*Sacha Bourgeois-Gironde and Éric Monnet*

**In the last twenty years**, we have seen the rapid emergence and growing influence of publications in economics that embrace the use of “natural experiments.” In this disciplinary context, natural experiments are defined as real situations that researchers do not control but that nonetheless have features that make them similar to laboratory experiments, meaning they can be used to identify causal relations. When proponents of this method began applying it to economic history, they elicited a particular kind of interdisciplinary interest—and lively debate. Natural experiments became the banner of a new historiographical movement whose practitioners were primarily economists intent on revealing the causality of historical and social phenomena, specifically the long-term economic effects of institutions. They called for using history not only to inform or decide the outcome of debates in economics but also, as explicitly stated in several prominent publications, to revisit major debates in the field of history and, in so doing, to give historians a lesson on method. Given the identities and influence of its members,

This article was translated from the French by Amy Jacobs-Colas and edited by Robin Emlein, Chloe Morgan, and Nicolas Barreyre.

\* An earlier version of this text was presented at two conferences, one entitled “Les expériences de pensée” (ENS ULM, 2014), the other “Histoire et causalité” (EHESS, 2015). Our thanks to the conference organizers and participants for their many comments and suggestions. Denis Cogneau, Sophie Cras, Claude Diebolt, Jacques Revel, and Alain Trannoy provided helpful, detailed remarks on earlier versions, and our discussions with Guillaume Calafat and François Keslair were extremely valuable. We alone are responsible for the interpretations and claims put forward here.

this school has sometimes been considered—by both proponents and critics—as a renewal of “the new economic history” (or cliometrics), which began imposing the use of economic models in historical study in the late 1950s.<sup>1</sup> In a survey of the current situation in economic history, Ran Abramitzky defined this approach as “modern economic history” and distinguished two main ways of using natural experiments: doing “economic history to test economic theory,” and doing “‘natural experiments’ to study the long-term impact of the past.”<sup>2</sup>

The most notable interdisciplinary publication in this field remains the anthology edited by Jared Diamond and James Robinson in 2010, which presents itself as both a manifesto and a progress report.<sup>3</sup> Despite its explicit title, the only texts with an epistemological dimension are the editors’ prologue and afterword. A number of chapters have become milestones in the study of causal effects in history; we will be discussing some of them here. Reactions to the book were as strong as its stated ambition: namely, to put forward an interdisciplinary scientific method for historical analysis modeled on the experimental sciences. Some historians saw this as a promise of scientificity and a revolution in our understanding of causality in history<sup>4</sup>; others concluded that the claims being made for this approach were excessive.<sup>5</sup>

Studies by the book’s contributors and the authors they cite—most of them renowned professors of economics in the United States, or their students—soon

1. Maurice Lévy-Leboyer, “La ‘New Economic History,’” *Annales ESC* 24, no. 5 (1969): 1035–69; Jean Heffer, “Une histoire scientifique. La nouvelle histoire économique,” *Annales ESC* 32, no. 4 (1977): 824–42; Claudia Goldin, “Cliometrics and the Nobel,” *Journal of Economic Perspectives* 9, no. 2 (1995): 191–208. According to several recent, more or less critical essays, the natural experiments movement in history represents either the continuation or the potential successor of new economic history. Those texts include Peter Temin, “The Rise and Fall of Economic History at MIT,” *History of Political Economy* 46, no. 1 (2014): 337–50; Temin, “Economic History and Economic Development: New Economic History in Retrospect and Prospect,” in *Handbook of Cliometrics*, ed. Claude Diebolt and Michael Hauptert (Berlin: Springer, 2016), 33–51; Francesco Boldizzoni, *The Poverty of Clío: Resurrecting Economic History* (Princeton: Princeton University Press, 2011). There has also been lively debate on the relevance of dubbing the many applications of natural experiments to African history “the new economic history of Africa.” For an introduction to this issue, see Morten Jerven, “A Clash of Disciplines? Economists and Historians Approaching the African Past,” *Economic History of Developing Regions* 26, no. 2 (2011): 111–24; Denis Cogneau, “The Economic History of Africa: Renaissance or False Dawn?” *Annales HSS (English Edition)* 71, no. 4 (2016): 539–56.

2. Ran Abramitzky, “Economics and the Modern Economic Historian,” *Journal of Economic History* 75, no. 4 (2015): 1240–51, here pp. 1245–47.

3. Jared Diamond and James A. Robinson, eds., *Natural Experiments of History* (Cambridge: Belknap Press of Harvard University Press, 2010).

4. Randolph Roth, “Scientific History and Experimental History,” *Journal of Interdisciplinary History* 43, no. 3 (2013): 443–58.

5. Joel Mokyr, review of Diamond and Robinson, *Natural Experiments of History*, *American Historical Review* 116, no. 3 (2011): 752–55. Mokyr concludes by stating that if the approach reined in its ambitions, it could in fact lead to better interdisciplinary use of comparative history, as advocated by Marc Bloch.

became highly successful and exerted a growing influence in economics journals, a virtually unprecedented situation for research using historical materials.<sup>6</sup> While the unity of the studies is obvious—indeed, they constitute a relatively autonomous sphere in history and economics publications—their diversity and evolutions should not be disregarded. As we shall see, many of these transformations reflect the fundamental epistemological problems this type of research has encountered in defining causality and in explaining how its statistical analyses are related to theory. Despite these variations, however, many current studies continue to share the same core queries and methods.

The aim of the present article is twofold. First, it approaches this research epistemologically in an attempt to understand why economists became interested in doing historical analysis through the lens of “natural experiments,” laying claim to a scientific approach that contrasts with the methods that they associated with research by historians. To proceed here we need to return to the definition of causality that has been widely applied in economics, particularly microeconomics, for the last twenty years at least. The empirical studies that follow this approach have become the new mainstream, and the specific definition of causality they use—one that is hardly self-evident or hegemonic in other fields—is based on two important postulates. One, it posits that logical causality and chronological (or historical) causality are juxtaposed; two, it aligns with an interventionist theory of causality. These two characteristics explain why economists extended the “natural experiments” method to historical questions, their assumption being that even though the “historical sciences” and “laboratory sciences” are distinct and subject to different constraints, they nonetheless share a common conception of causality. The two postulates also enable us to understand why this approach developed primarily within the framework of the neo-institutionalist paradigm, which lends itself particularly well to an interventionist theory of causality.<sup>7</sup> Having a firm grasp of these characteristics and the connections between them will allow us to see why this way of proceeding has remained relatively separate within the social sciences despite its success in economics—in fact, this body of work hardly interacts at all with other historiographical approaches—and why the vast majority of “natural experiment” studies using historical data have been done within the neo-institutionalist framework when there is no a priori reason to assume that causal analysis is restricted to that framework. Paradoxically, the interventionist conception of causality amounts to rehabilitating the role of events in history—as exogenous causes, comparable to “manipulations” of the course of history—within an analytic framework that favors a long-term approach (the study of an event’s long-term effects).

6. Abramitzky, “Economics.”

7. This obviously does not mean that any economics study based on this type of causality necessarily fits into the neo-institutionalist paradigm. Here we detail the specific problems involved in applying *to history* natural experiment methods that draw on an underlying theory of institutions, and explain how these problems differ from the ones raised in criticisms of this type of causal approach applied in other fields.

In the second section we critique this way of proceeding, bringing to bear both external and internal analyses and highlighting the logical aporias in the notion of natural experiment, contradictions brought to light, as we shall see, by Judea Pearl's different notion of causality. The difficulties that the dominant definition of causality in economics and the notion of natural experiment encounter when it comes to taking into account temporality and defining counter-factuality considerably restrict the potential for applying this approach to historical phenomena. Other conceptions are possible and deserve to be explored. Indeed, our aim here is to contribute to more general, interdisciplinary reflection on causality in history and its relation to temporality. Scholarship in history based on the natural experiments approach can, at the very least, be credited with defining and implementing a relatively homogeneous method for analyzing the long term, one founded entirely on the notions of causality and institutions—and on singular definitions of them. Probing these presuppositions and definitions is therefore a precondition if we are to compare this approach critically to other possible methods and point out its shortcomings. In their polemical work, Jo Guldi and David Armitage invite historians to conjointly reappropriate the notions of long term and causality (insisting on the multiplicity of causes),<sup>8</sup> while mistakenly asserting that economists have neglected the problem of causality.

## Old and New Economic History: The Effects of the Empirical Turn

The body of publications that claim to contribute something new to economic history on the basis of methods borrowed from the experimental sciences and the identification of causality should not be confused with what is commonly called new economic history (and is in fact already old). The difference between new economic history, which emerged in the 1950s and 1960s, and the more recent studies discussed here reflects a progression in economics towards an empirical approach whose main objective is to identify causal relations, a distinct change from formerly dominant theoretical models. What is now termed “applied economics” has won out for the time being, its hegemony achieved through the definition of new criteria of scientificity now applied in academic journals, including when the subject material is historical.<sup>9</sup> The priority is no longer to construct and test theories but rather to focus

8. Jo Guldi and David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014), 110; and the dossier entitled “Debating the *Longue Durée*,” *Annales HSS (English Edition)* 70, no. 2 (2015): 215–303.

9. Michel De Vroey and Luca Pensieroso, “The Rise of a Mainstream in Economics” (IRES discussion paper no. 26, Institute for Economic and Social Research, Université catholique de Louvain, 2016), 2–27; Matthew T. Panhans and John D. Singleton, “The Empirical Economist’s Toolkit: From Models to Methods” (working paper no. 3, Center for the History of Political Economy, Duke University, Durham, 2015), 1–27.

on a set of methods allowing researchers to measure the causal effect of a given economic policy or event.<sup>10</sup> The resulting association between natural experiment and causality is particularly well explained in an influential text by the economists Joshua Angrist and Jörn-Steffen Pischke, in which they argue that the method amounts to a genuine scientific revolution: “Design-based studies are distinguished by their *prima facie* credibility and by the attention investigators devote to making both an institutional and a data-driven case for causality.”<sup>11</sup>

The widespread adoption of the term “natural experiment” reflects the assumption that a laboratory, or randomly controlled, experiment is the ideal scientific proceeding. Its application to history is only one example of this approach, which has come to dominate in applied microeconomics (for issues such as the labor market and education) in reaction to the use of statistical methods that confused cause and correlation and did not explain the foundations of their analyses. In the *Annales* in 1977 Jean Heffer said of the then-new economic history: “The crucial point is theory.”<sup>12</sup> The same cannot be said today of the recent studies in economic history that make use of natural experiments; on the contrary, nowadays the crucial point is causality.

But the difference between new economic history and the current trend towards using natural experiments in economic history is more than methodological; it is also reflected in distinct outlets for publication.<sup>13</sup> The two groups’ respective relations to macroeconomics and the long term are also very different. In the 1950s and 1960s, new economic historians borrowed mainly from theoretical macroeconomic models (often general equilibrium models), which they then used to quantify and clarify historical processes such as the calculation of returns on labor and capital, national accounting, the monetary theory of fluctuations, and so forth. In direct contrast—and paradoxically—the natural experiments approach, although mainly applied to history on a macroeconomic scale, derives its empirical methods foremost from microeconomics. The natural experiments approach to economic history tries to have it both ways, seeking to import methods of causal identification into both economic history and macroeconomics.<sup>14</sup>

10. Joshua D. Angrist and Alan B. Krueger, “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments,” *Journal of Economic Perspectives* 15, no. 4 (2001): 69–85; Joshua D. Angrist and Jörn-Steffen Pischke, “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics,” *Journal of Economic Perspectives* 24, no. 2 (2010): 3–30. For a discussion of these methods as applied in political science, see Allison J. Sovey and Donald P. Green, “Instrumental Variables Estimation in Political Science: A Readers’ Guide,” *American Journal of Political Science* 55, no. 1 (2011): 188–200; Jasjeet S. Sekhon and Rocío Titiunik, “When Natural Experiments are Neither Natural nor Experiments,” *American Political Science Review* 106, no. 1 (2012): 35–57.

11. Angrist and Pischke, “The Credibility Revolution,” 5.

12. Heffer, “Une histoire scientifique,” 824–25.

13. Abramitzky, “Economics.”

14. In Angrist and Pischke’s manifesto “The Credibility Revolution in Empirical Economics,” the possible applications of this method to macroeconomics are largely drawn from studies using natural experiments in economic history.

We can, however, identify a bridge between one component of new economic history and many of the studies claiming to conduct natural experiments in history, namely neo-institutionalism.<sup>15</sup> The analyses cited above, particularly those in the volume edited by Diamond and Robinson, and the ones we examine below, are directly descended from the work of Douglass North and Avner Grief in that they posit institutions as drivers of economic development and seek to compare different institutions' economic effects. They share three characteristics with neo-institutionalism, characteristics that also sharply differentiate this kind of study from the "old" institutionalism of Max Weber, Thorsten Veblen, and Karl Polanyi (itself influenced by the German historical school). First, in both neo-institutionalism and natural experiment analyses, institutions are viewed as equilibria resulting from incentives and coordination procedures that produce behavioral regularities and particular economic effects. Institutions are what enable markets to operate. Contrary to the earlier institutionalist theories, which considered the market itself as an institution, today's approach makes a fundamental distinction between institutions and markets. Second, like neo-institutionalism, natural experiment analysis considers that similar types of institutions can be observed—and described—in different contexts.<sup>16</sup> Third, both currents postulate that it is possible to identify certain types of institutions that lead to greater economic growth and to act on those institutions to improve economic performance. A major difference between the "old" institutionalism and "neo-institutionalism" has to do with the fact that the latter seeks to identify types of institutions that are similar and explain why they are able to emerge in relatively different contexts. It compares institutions over the long term and on a large geographical scale, even though institutional mechanisms and balances are explained in microeconomic terms—by way of game theory, for example.<sup>17</sup> The original institutionalists, on the other hand,

15. This connection is studied by Boldizzoni in *The Poverty of Clio*.

16. In *Why Nations Fail: The Origins of Power, Prosperity, and Poverty* (New York: Crown Business, 2013), Daron Acemoglu and James A. Robinson identify two types of institutions said to be observable in different periods and regions: "extractive" and "inclusive."

17. For a comprehensive introduction to neo-institutionalist method and how it is applied in economic history, see Avner Greif, *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade* (Cambridge: Cambridge University Press, 2006); and Robert Boyer's critical assessment, "Historiens et économistes face à l'émergence des institutions du marché," *Annales HSS* 64, no. 3 (2009): 665–93. See also Guillaume Calafat, "Familles, réseaux et confiance dans l'économie de l'époque moderne. Diasporas marchandes et commerce interculturel," *Annales HSS* 66, no. 2 (2011): 513–31. In *Why Nations Fail*, Acemoglu and Robinson present both their neo-institutionalist notion and the distinction between extractive and inclusive institutions, a distinction which builds on that made by Douglass North in his last writings; see especially Douglass C. North, John Joseph Wallis, and Barry R. Weingast, *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History* (Cambridge: Cambridge University Press, 2009). For a critical discussion of the filiation between North, Wallis, and Weingast and their references, see Martin Daunton, "Rationality and Institutions: Reflections on Douglass North," *Structural Change and Economic Dynamics* 21, no. 2 (2010): 147–56.



emphasized the historical development of singular institutional forms over time. We must therefore explain how neo-institutionalism, by identifying the causality operative in natural experiments, has produced a new type of historical analysis.

## Causality, History, and Intervention

What concept of causality do proponents of the natural experiments method apply to history? In this context, the effect of variable A on variable B is thought of as causal if *one*, A determines B and is not determined by B (the relation must be one-way and nonsimultaneous), and *two*, A's effect on B is direct, that is, it must be observed when other parameters remain constant (all else being equal); there can be no "omitted variable" that could jointly influence A and B. When both these conditions are met, it can be said that A is exogenous to B, which is equivalent to claiming that A is the cause of B. The notion of "exogeneity" is central to this notion of causality.<sup>18</sup> In this idealized framework, an exogenous intervention (the administering of a substance, for example) affects a group whose characteristics are identical to those of a control group unaffected by the intervention (it was not given the substance). Researchers then try either to reproduce this idealized framework as nearly as possible (a controlled experiment) or to find a real-world situation that can be considered close to it (a natural experiment).<sup>19</sup> They must therefore focus closely on the research design of the experiment.<sup>20</sup> In the case of natural experiments, researchers themselves cannot play a designing role but must instead argue persuasively that reality resembles a laboratory research design in certain respects. This in turn means they must prove two things: first, that the real event in question, which they conceive of as an intervention, is indeed "exogenous" to the thing it affects (it is precisely that impact that the researchers intend to measure); second, that there is a "control group" with characteristics identical to or closely resembling those of the group affected by the event or, failing that, that all differences between the two groups can be controlled for in statistical estimations of the expected effect (the "all else being equal" reasoning). If researchers cannot identify a situation in which the same event has different repercussions on identical groups, they can turn to the "instrumental variables" method, wherein causality is identified using a variable that affects the factor whose effect is to be measured—the assumed cause—while not being determined by the effect to be observed—the consequence—and that can only affect the consequence by way of the assumed cause. Though, technically

18. Technically, exogeneity is ensured when the independent variable in an econometric model is not correlated with the residuals. In addition to the two conditions stated above, measurement errors in the independent variable can also lead to correlation with the residual.

19. An experiment is understood to be "controlled" when the researcher designs it and intervenes. An example might be comparing a group of students who benefited from an education reform with another group who were not exposed to it.

20. Angrist and Pischke, "The Credibility Revolution."

speaking, the instrumental variables method does not follow the same protocols as a controlled experiment, it is based on an equivalent definition of causality and is therefore associated with the natural experiments method.<sup>21</sup>

Clearly, this type of causality is defined in reference to an idealized laboratory protocol, in which a voluntary external intervention triggers an event that necessarily occurs after it, independently of movement in other parameters, and of which it can be said that it would not have taken place without the intervention. In this approach, conducting analysis outside the laboratory is thought of as a “fallback” method, the whole point of the scientific proceeding being to adhere as closely as possible to the framework posited as ideal.<sup>22</sup> The rhetoric of comparing economic analysis to laboratory practice is everywhere.<sup>23</sup> In their prologue, Diamond and Robinson justify turning to history precisely because it is impossible to construct “historical” laboratory experiments. Sciences concerned with the past are at a disadvantage compared to natural sciences because, by nature, the past cannot be manipulated or made into a laboratory experiment: “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.”<sup>24</sup> The aim of the historical sciences, then, should be to use history to identify natural experiments and so to establish “chains of cause and effect.”<sup>25</sup> It is this temporal (or chronological) notion of causality that underlies econometric methods, and particularly the idealized research design of this branch of analysis.<sup>26</sup> Diamond and Robinson are implicitly

21. *Ibid.* See also Mark R. Rosenzweig and Kenneth I. Wolpin, “Natural ‘Natural Experiments’ in Economics,” *Journal of Economic Literature* 38, no. 4 (2000): 827–74. For an epistemological discussion of instrumental variables, see Julian Reiss, “Causal Instrumental Variables and Interventions,” *Philosophy of Science* 72, no. 5 (2005): 964–76.

22. Angrist and Pischke, “The Credibility Revolution.”

23. It goes without saying that total control of outside parameters is seldom possible, even in so-called “randomized controlled” experiments: see Agnès Labrousse, “Learning From Randomized Controlled Experiments: The Narrative of Scientificity, Practical Complications, Historical Experience,” *La vie des idées* 2016, <https://booksandideas.net/Learning-from-Randomized-Controlled-Experiments.html>; Angus Deaton and Nancy Cartwright, “Understanding and Misunderstanding Randomized Controlled Trials” (NBER working paper no. 22595, National Bureau of Economic Research, Cambridge, 2016). Experimental economics makes extensive use of laboratory experiments with individuals to test economic theories about behavior but, paradoxically, seldom uses the notion of causality as employed by applied economics in studies on economic policy realized outside of a laboratory setting. More clarifications are needed on the sources of the notion that laboratory experiments are the ideal study situation; here we simply analyze the presuppositions underlying the definition of causality associated with the laboratory reference.

24. Diamond and Robinson, *Natural Experiments*, 1.

25. *Ibid.*

26. Adopting a temporal notion of causality (cause precedes effect) does not mean assuming that researchers can fully account for temporality itself (the succession of events and whole chains of cause and effect). See the second half of the present article.

defending the idea that logical and historical causality are juxtaposed. This classic, fundamental postulate of causality analysis, which amounts to claiming that causes necessarily precede their effects, should be emphasized. From this first, highly intuitive postulate and its implicit acceptance in natural experiment methodology directly proceeds a juxtaposition between the flow of history (from past to present) and causal links (causes precede their effects). History offers the vital condition of time's arrow, an asymmetrical temporal direction of cause towards effect.<sup>27</sup> Given that variables of interest are situated in time, it becomes tempting to speak of historical causality rather than mere statistical causality. Economic models, meanwhile, are often atemporal and are based solely on the idea of cause as the logical antecedent of effect.<sup>28</sup>

The second postulate of natural experiment analysis holds that, while we cannot manipulate the past, the past offers examples of manipulation. As there is no demiurge to change the course of history, the name given to this ideal experimenter is "nature" (a concept which cannot be reduced to biological or climatic phenomena). This definition amounts to an interventionist or "manipulativist" theory of causality<sup>29</sup>: reality can be "manipulated," intervened in, its course changed. This assumption—that we can identify a sufficiently specific, independent, manipulable variable among the mass of historical material—directly contradicts several alternative definitions of causality, as we shall see in the second half of this article. Here let us just point out that it raises a certain problem of principle regarding whether it is possible to clearly distinguish in history between what is a matter of intervention—a break in the course of history, which would have continued on as before if there had not been that putative intervention—and the course of history itself, of which the intervention is wholly a part. In general, interventionist approaches to causality are reductionist in that they ascribe a predominant role to human action, assuming that action—free, not entirely determined by the course of history itself—can somehow remove or disengage itself from history. Other theorists of this analytic approach in philosophy refuse to grant this kind of metaphysical exception to human action.<sup>30</sup> The question then arises of how such postulates could support historical analysis.

On the one hand, the definition of causality that has come to dominate in economics, to the point of becoming the new standard of scientificity, rests on an equivalence between temporal and logical causality. On the other hand, this

27. For a deeper analysis of this postulate of the temporal asymmetry of causes and effects and the conceptual possibility of reverse causality, see Bourgeois-Gironde, *Temps et causalité* (Paris: Presses universitaires de France, 2002).

28. The paradigmatic example is price determination in the neoclassical model.

29. For a particularly well-formulated version of this definition, see James Woodward, *Making Things Happen: A Theory of Causal Explanation* (Oxford: Oxford University Press, 2005). For a discussion of instrumental variables in that framework, see Reiss, "Causal Instrumental Variables."

30. Judea Pearl, "Causal Inference," in "Causality: Objectives and Assessment," ed. Isabelle Guyon, Dominik Janzing, and Bernhard Schölkopf, *Proceedings of Machine Learning Research*, vol. 6 (2010): 39–58.

type of causality is of the interventionist (or manipulativist) variety, as attested by the omnipresence of references to ideal laboratory experiments and to the role attributed to human action in triggering causal chains. In our view, these two characteristics explain why and how economists have applied their definition of causality to history. As Diamond and Robinson explicitly state, once it has been acknowledged that laboratory experiments are impossible in sciences “concerned with the past,” then historical analysis becomes a sort of natural extension, by default, of this temporal, interventionist notion of causality. This then gives rise to the distinction between historical and laboratory sciences, which is not reducible, for instance, to that between the social and human sciences and the natural sciences. The interventionist type of causality is particularly compatible with the neo-institutionalist paradigm in that they share a reductionist approach to human action and a decontextualized definition of institutions. Their similarity is evident in the common use of the terms “design-based studies” and “designed institutions.”<sup>31</sup> In neo-institutionalist theory, an institution’s effect on social behavior and economic markets is comparable to that of an experimenter intervening in a laboratory setting. In both cases, researchers hypothesize that manipulation (design) is possible and that interventions are independent of the course of history. This epistemological proximity is utilized extensively to study the long-term effects of institutions by harnessing the notions of causality and natural experiment.

## Generalizing

For proponents of natural experiments, history admits causal analysis and therefore allows for the making of generalizations. In their afterword, Diamond and Robinson stress this objective, which, as they see it, stands in opposition to a traditional historical approach focused on singular cases.<sup>32</sup> But what is the nature of the generalizations that researchers should make? Should they be inductive or deductive? An inductive generalization sees history as an accumulation of causal relations. The criticism generally directed at this type of generalization is relatively intuitive and has been leveled at many microeconomic studies: the experiments they present are too heterogeneous and have an indeterminate relation to theory, which reduces their “external validity.”<sup>33</sup> Applied to history, the criticism can be formulated as follows: In order for generalization to be possible, historical events

31. As can be seen in the articles discussed here, specifically in those by Angrist and Pischke (“design-based studies”) and Acemoglu et al. (“designed institutional change”). The more general notion of “research design” is widely used to describe the methods of applied economics, while “institutional design” is recurrent in the work of neo-institutionalists.

32. Diamond and Robinson, *Natural Experiments*, 271–74.

33. Nancy Cartwright, “Are RCTs the Gold Standard?” *BioSocieties* 2, no. 1 (2007): 11–20; Deaton and Cartwright, “Understanding and Misunderstanding.”

must be similar; but asserting their similarity amounts to denying the historical contexts of the events in question.

Manifestoes in favor of scientificity in the social sciences have long stressed the need to produce general statements on the basis of unique cases, and this call is closely linked to the pursuit of causality. As François Simiand so clearly put it in an early twentieth-century series of texts on scientific methods in the discipline of history, there is no causality without law because we can only identify a cause if we can assert that it always produces the same effects.<sup>34</sup> In this understanding of causality, the reference to a theory, to a general proposition, is indispensable if we are to distinguish a cause from a mere condition; that is, from the fact that, all else being equal, one event preceded another. The ability to generalize is thus intrinsically linked to causal analysis. In asserting that there can only be a cause if there is a law, Simiand was much clearer on this point than Diamond and Robinson. In their volume, the status of natural experiments is ultimately likened to that of comparisons (particularly in their afterword), despite the fact that the fundamental opposition between particular cases and general statements remains.<sup>35</sup> They therefore equate causal analysis, comparative analysis, and the ability to generalize without considering that comparative analysis can be non-causal or founded on a different notion of causality, and without clarifying the relationship that causal analysis must have to theory in order to produce general statements.<sup>36</sup> In applied economics and the evaluation of public policy, researchers can defend the radical empiricism of their approach by invoking either the fact that their work is not intended to be generalist in scope and does not have “external validity” or the notion that generalizations can only be constructed gradually through an accumulation of knowledge obtained by reproducing similar research protocols (as in medicine). For natural experiments in history, however, there is relatively little reason to hope that conclusions can be generalized on the basis of a single case or that similar protocols can be reproduced. In the absence of a law, generalizations can only be made by linking a singular process to a more general historical one. In fact, in most studies that use the natural experiments method, institutionalist theory serves as the basis for forming generalizations.

34. François Simiand, “Méthode historique et science sociale. Étude critique d’après les ouvrages récents de M. Lacombe et de M. Seignobos,” *Revue de synthèse historique* 16 (1903): 1–22; Simiand, “La causalité en histoire,” *Bulletin de la Société française de philosophie* 6 (1906): 245–90. On the context of Simiand’s position on causality and the debates it sparked, see Jacques Revel, “Histoire et sciences sociales. Lectures d’un débat français autour de 1900,” *Mil neuf cent. Revue d’histoire intellectuelle* 25, no. 1 (2007): 101–26.

35. Diamond and Robinson, *Natural Experiments*, 271–74. From the outset, the editors explain that they think of “natural experiments” as similar to the “comparative method,” but define them only in reference to laboratory experiments, not to other comparative strategies in the discipline of history.

36. *Ibid.* So little is offered here on the status of generalization, on relations to theory, and therefore on the question of laws that we do not know, for instance, if the authors would go so far as to conclude that “laws of history” exist and that causal analysis is intended to reveal them.

## The Long Term and Causality

Let us now turn our attention to one study that uses the French Revolution as a natural experiment. Without delving here into the data used or how the study relates to the historiography of the Revolution, we will show how it illustrates the methods and difficulties outlined above. In this case at least, as we shall see, the authors were not able to establish a clear tie between their natural experiment, history, and a broader theory. Published initially in the volume by Diamond and Robinson before being reworked for the *American Economic Review*, this study is one of the main references for this type of research in economic history.<sup>37</sup> The title—“From Ancien Régime to Capitalism: The French Revolution as a Natural Experiment”—conveys the authors’ intention to use the French Revolution to study the shift from the ancien régime to a capitalist system. In their first version, they criticized historians for misunderstanding causality and affirmed that the identification of causal relations is essential to the formulation of general conclusions:

*Importantly, we argued that there was something special about this historical episode; it can be thought of as a natural experiment that gives us the possibility of saying something much more precise about causal factors than is typically the case in historical or social studies. ... History is full of such potential experiments; it is just that historians have not yet thought of them in these terms. We believe that exploiting these experiments in a systematic way will greatly improve our understanding of the important forces that have driven long-run processes of historical, social, political and economic change.*<sup>38</sup>

For the authors, the exogeneity of the French invasion of Prussia and the reforms that followed, imposed on Prussian territories purely at random without taking into account the territories’ economic particularities, justifies their use of the term “natural experiment.” In this understanding, the conquered Prussian regions serve as the “treatment” group and the conquering French armies the “control group.” Much of the article is devoted to demonstrating that the subjugated territories were not invaded because they were richer than others or showed greater potential for economic growth. By defining this situation as an experimental research design, the authors seek to establish the economic causal effects of the invasion, comparing the economic situations of invaded and non-invaded Prussian territories several decades later. In their initial version, they availed themselves of these methods to measure the economic impact of abolishing ancien régime institutions, thereby testing Friedrich Engels’s hypothesis that territories invaded by Napoleon

37. Daron Acemoglu et al., “From Ancien Régime to Capitalism: The French Revolution as a Natural Experiment,” in Diamond and Robinson, *Natural Experiments*, 221–56; Daron Acemoglu et al., “The Consequences of Radical Reform: The French Revolution,” *American Economic Review* 101, no. 7 (2011): 3286–307.

38. Acemoglu et al., “From Ancien Régime to Capitalism,” 249–50.

experienced greater economic growth due to the fact that existing elites lost their power. Ultimately they observed that in the second half of the nineteenth century—that is, once enough time had gone by, in their estimation, for the invasion’s full effects on institutions to be felt—the population grew more rapidly in the conquered territories. Since population growth is understood as a sign of economic development, they concluded that the Napoleonic conquests had a beneficial effect.

The successive versions of this text provide a glimpse of different interpretations of the theory being tested in this natural experiment (and therefore of different interpretations of the causal relation to be identified and different possible general conclusions). They also reveal the authors’ doubts about whether it was necessary to refer to a theory in order to determine causal relation. In the second version, published in an economics journal and taking into account remarks by external reviewers, the authors drew broader political and social conclusions that were also more firmly rooted in the neo-institutionalist paradigm:

*The French Revolution is a clear example of a large-scale, radical, and “designed” institutional change. In this light, our findings support the centrality of institutional differences for comparative economic development. More important, the results are inconsistent with the view that externally imposed, radical, and “Big Bang” style reforms can never be successful.<sup>39</sup>*

The article speaks little of theory, except to note that the presence of an oligarchy (and of ancien régime institutions in general) stymied growth due to the existence of significant economic rents and the appropriation of resources by an elite. The authors refer in this second version to a model in which that mechanism is developed more formally and analytically, although their aim is not to test it.<sup>40</sup> In fact, the model they discuss serves as a general institutionalist framework enabling the authors to establish a tie between the French invasion, institutional reforms, and, ultimately, growth. Their purpose is to show that the link between observations about different moments in time is indeed causal. By relying on this underlying institutionalist theory, the authors can claim that what they have observed is causality and not a fallacious correlation between two temporally distant phenomena. To construct their counterfactual and justify their design, they must, at the very least, assume a relation between institutional events and population growth.<sup>41</sup>

39. Acemoglu et al., “The Consequences of Radical Reform,” 3303.

40. Daron Acemoglu, “Oligarchic versus Democratic Societies,” *Journal of the European Economic Association* 6, no. 1 (2008): 1–44.

41. On applying similar methods and reasoning in other contexts, see Daron Acemoglu, Tarek A. Hassan, and James A. Robinson, “Social Structure and Development: A Legacy of the Holocaust in Russia,” *Quarterly Journal of Economics* 126, no. 2 (2010): 895–946; Sara Lowes et al., “The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom,” *Econometrica* 85, no. 4 (2017): 1065–91. In both these cases, it is again institutionalist theory that links actions, rules, and human behavior and allows for interpreting as causal a correlation between two observations separated by several centuries.

## Testing Theories or Establishing Long-Term Causality?

The ambiguity of this example sheds light on how the use of natural experiment methods in economic history has ultimately led to two distinct types of investigations, as noted by Abramitzky in the text mentioned above. One type directly uses history to test theories,<sup>42</sup> formulating theory-driven predictions and probing history for situations that can be used to test those theories, instances where the author's observations correspond to those predicted by the theoretical model. This method, largely founded on Karl Popper's Falsification Principle, is well established in economics, though it has obviously not escaped criticism.<sup>43</sup> Here we need only say that this approach is not new. Moreover, it exists independently of natural and laboratory experiments, since Popper's falsification method assumes no form of causality. Economists have always used historical data to test economic theories, without necessarily trying to broaden our understanding of economic history and thereby to rival the work of historians. The natural experiments approach is simply seen as one of many possibilities—although its popularity has significantly increased in recent years. It is fairly standard, for example, to use long price series to test financial theories, without necessarily invoking causality.<sup>44</sup>

The other type of investigation proceeds in the opposite direction, adopting the natural experiments approach primarily to improve our understanding of economic processes in history instead of limiting its application to the testing of theories.<sup>45</sup> In the article on the French Revolution, for example, the analysis is firmly focused on the impact of the Revolution on ancien régime institutions.

42. Réka Juhász, "Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade" (Centre for Economic Performance discussion paper no. 1322, London School of Economics, 2014); Peter Koudijs, "The Boats That Did Not Sail: Asset Price Volatility in a Natural Experiment," *Journal of Finance* 71, no. 3 (2016): 1185–226. The criticisms presented in the second half of the present article also apply to this type of study.

43. See Daniel M. Hausman, ed., *The Philosophy of Economics: An Anthology* (Cambridge: Cambridge University Press, 1984; repr. 1994); Julian Reiss, *Error in Economics: Towards a More Evidence-Based Methodology* (London: Routledge, 2008; repr. 2016).

44. Two renowned articles come to mind here: Rajnish Mehra and Edward C. Prescott, "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15, no. 2 (1985): 145–61; Karl E. Case and Robert J. Shiller, "The Efficiency of the Market for Single-Family Homes," *American Economic Review* 79, no. 1 (1989): 125–37.

45. In addition to the aforementioned texts by Diamond and Robinson, and Acemoglu and Robinson's *Why Nations Fail*, the following articles are particularly clear on this point: Nathan Nunn, "The Importance of History for Economic Development," *Annual Review of Economics* 1, no. 1 (2009): 65–92; James Fenske, "The Causal History of Africa: A Response to Hopkins," *Economic History of Developing Regions* 25, no. 2 (2010): 177–212.



A number of other articles using similar methods focus on the effects of colonialism,<sup>46</sup> with the principal objective of restoring a role to institutions in explaining economic development in Africa. Those studies make no reference to Africanist research, and their authors are in open disagreement with a body of earlier economic research that they see as misguidedly obsessed with geographic and environmental factors. They argue that in those places where colonial institutions did the greatest harm to local economies (where they were the most “extractive”), later growth was weaker. To measure institutional characteristics exogenous to growth, the authors hypothesize that these characteristics were determined by the mortality of colonizers, linked in turn to climate. Since a favorite theme in this literature is the long-term effects of colonialism, historians specialized in African economic history have begun debating its findings, occasionally applauding its boldness but also criticizing its methodological shortcomings.<sup>47</sup> This literature can also be compared to studies in which researchers use the natural experiments method to focus on the economic effects of Protestantism.<sup>48</sup> In neither case do they try to reduce history to a field for testing economic theories, endeavoring instead to build a general theory of historical development.

The theory these studies propose is different from the notion of causality found in Weber’s own work, which is necessarily referenced in any literature on the effects of Protestantism. Weber not only envisioned multi-causal factors (that is, a single phenomenon explained by several causes) but also, more importantly,

46. Daron Acemoglu, Simon Johnson, and James A. Robinson, “The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review* 91, no. 5 (2001): 1369–401; Acemoglu, Johnson, and Robinson, “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution,” *Quarterly Journal of Economics* 117, no. 4 (2002): 1231–94.

47. Gareth Austin, “The ‘Reversal of Fortune’ Thesis and the Compression of History: Perspectives from African and Comparative Economic History,” *Journal of International Development* 20, no. 8 (2008): 996–1027; Antony G. Hopkins, “The New Economic History of Africa,” *Journal of African History* 50, no. 2 (2009): 155–77; Hopkins, “Causes and Confusions in African History,” *Economic History of Developing Regions* 26, no. 2 (2011): 107–10; Jerven, “A Clash of Disciplines.”

48. For a recent review of this literature, see Sascha O. Becker, Steven Pfaff, and Jared Rubin, “Causes and Consequences of the Protestant Reformation,” *Explorations in Economic History* 62, no. 3 (2016): 1–25. The authors define the research procedures in this new field as identifying causality using econometric procedures and estimating long-term effects from an institutionalist perspective. An exception is Davide Cantoni’s “test” of the relation Weber established between the Protestant ethic and development; see Cantoni, “The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands,” *Journal of the European Economic Association* 13, no. 4 (2015): 561–98. Cantoni finds no positive effect on economic growth in regions that converted to Protestantism and concludes that Weber’s theory has been invalidated. Although, strictly speaking, he only assesses links between the differentiated impacts of the Peace of Augsburg (presented as the source of a natural experiment) and population growth in particular German cities, the initial ambition of producing a generalization leads the author to present his work as an econometric test of a major theory. He is ultimately compelled to downsize that claim.

unique causal relations (wherein a particular causality chain is associated with a particular moment in a particular context and cannot be reproduced in a different context). As he saw it, the emergence of capitalism was not due to the Protestant ethic alone but to how that ethic dovetailed with the structure of northern European medieval cities and the development of specific accounting systems. He favored a causal—in the sense of counterfactual—approach to the relationship between religion and economic development and constructed his arguments through comparisons of Europe to China and India; his line of argument did not require him to specify whether the different regions shared any characteristics other than religion. The point was to understand a singular event—the advent of capitalism in Europe—and to describe the unique combination of factors that facilitated its occurrence. Moreover, Weber emphasized the difficulty of observing any immutable tie between Protestant asceticism and capitalism (thus questioning the temporality of that relationship) since, as he put it, the development of capitalism had eradicated the religious foundations within Protestantism that had fostered its original expansion.<sup>49</sup>

At the turn of the twenty-first century, as interest in growth models based on production factors declined, natural experiment studies of economic history offered an answer to economists' new or renewed questions in the form of causal analyses of the effects of colonialism and Protestantism (made possible by the availability of data produced by colonial administrations and by extensive German demographic data). Combining the neo-institutionalist framework with references to Robert Putnam's theories of social capital, these studies posited that economic divergence was due to more or less effective fits between "institutions," "culture," "trust," and markets.<sup>50</sup> It is difficult at this stage to imagine long-term causality studies conducted outside the general framework offered by this paradigm, whose influence on American economists and political scientists has grown steadily since the late 1990s. Their work has advanced a definition of

49. Max Weber, *The Protestant Ethic and the Spirit of Capitalism* [1906], trans. Talcott Parsons (London: Routledge, 1930; repr. 2005), 124: "Since asceticism undertook to remodel the world and to work out its ideals in the world, material goods have gained an increasing and finally inexorable power over the lives of men as at no previous period in history. Today the spirit of religious asceticism—whether finally, who knows?—has escaped from the cage. But victorious capitalism, since it rests on mechanical foundations, needs its support no longer." To our knowledge, this crucial comment by Weber does not seem to have had any resonance in economic studies of the long-term link between the Protestant religion and economic growth.

50. Greif, *Institutions*; Robert D. Putnam, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton: Princeton University Press, 1993). For an introduction to how economists have appropriated these two schools of thought, see Guido Tabellini, "Presidential Address: Institutions and Culture," *Journal of the European Economic Association* 6, no. 3 (2008): 255–94. For a critique of how economists have used Putnam's work and the historical determinism that follows from it, see Nicolas Delalande, "Is a History of Trust Possible? Remarks on the Historic Imagination of Two Economists" [2008], *La vie des idées*, 2011, <https://booksandideas.net/Is-a-History-of-Trust-Possible.html>.

the long term based entirely on the notions of causality and institutions, while paradoxically placing “events” seen as exogenous at the center of their analysis. Readers not familiar with this field of research cannot help but be struck by the fact that the application of natural experiments to history is essentially limited to a small number of subjects—notably religion and colonialism—with high ideological stakes, especially when it comes to debating the economic superiority of certain institutional configurations. It is therefore all the more important to examine the underlying assumptions of these studies—in other words, to attempt to grasp what they are saying about human action and the meaning of history.<sup>51</sup>

## Limits and Criticism

This second part of our article presents the main methodological criticisms leveled at approaches that involve identifying long-term causality in history. These critiques have led several economists and historians to doubt the scientific utility of the entire approach or to denounce its ideological assumptions.<sup>52</sup> Some assessments have been neatly formulated by skeptical authors, while others have emerged through clashes with other conceptions of causality used in philosophy or the social sciences. Our aim here is to encourage historians and economists to take up these epistemological remarks and engage in a debate on the different uses of causality in economic history. The remarks in question mainly concern how causal methods drawn from natural experiments have been applied in the field of history; there is no need to restate more general criticisms of natural or controlled experiments that expose the lack of external validity and point out that the theoretical conditions required to identify causality are almost never verified.<sup>53</sup> Such criticisms are “internal” to the method and apply to the use of natural experiments in a range of contexts—the assessment of public policy, for example. They are fundamental and already widely recognized, although there is a tendency to ignore them when drawing political conclusions from research. The question that interests us here is specific to history: on the one hand because a considerable amount of time elapses between the event considered exogenous in the “research design” of the experiment and the observed variable of interest; and on the other hand because an interventionist definition of causality relies on hypotheses that imply a particular vision of historical development.

51. Our way of proceeding here resembles that of Quentin Deluermoz and Pierre Singaravélou in *Pour une histoire des possibles* (Paris: Éd. du Seuil, 2016), 219. These authors call for examining the political uses of counterfactuals in history and criticize the claims to scientificity of mechanical conceptions of causality that fail to apply counterfactual reasoning to their own “analytical framework.”

52. In addition to the references cited above, see Boldizzoni, *The Poverty of Clio*.

53. Rosenzweig and Wolpin, “Natural ‘Natural Experiments’”; Cartwright, “Are RCTs the Gold Standard?”; Deaton and Cartwright, “Understanding and Misunderstanding”; Sekhon and Titiunik, “When Natural Experiments.”

## A Retrospective Illusion and Some Limitations

The first type of criticism does not necessarily target the natural experiments method itself but expresses concern about its ambition to impose new criteria of scientificity on economic history or to unify the social sciences around a tightly circumscribed definition of causality. In its initial version, this critique consisted simply in observing that there are too few historical cases that resemble natural experiments for this approach to become the norm. If it did manage to win broad acceptance, it would drastically limit the subjects that could be legitimately investigated, thereby eventually subordinating the intrinsic interest of historical phenomena to the use of a single, specific method.<sup>54</sup> This is the substance of Naomi Lamoreaux's appeal to economists interested in long-term causal effects to ensure that their method does not skew their perception of the past. Her critique bears specifically on the strong neo-institutionalism associated with research on causal effects over the long term. It highlights the risk of retrospective determinism that follows from the hypothesis—often implicit, as we have seen—that institutions have observable, unique, and lasting effects on behavior. The clearest danger is to ascribe unwarranted, anachronistic importance to institutional forms perceived from the present—retrospectively, that is—as salient.<sup>55</sup> This type of critique amounts to denouncing the teleological bias behind the natural experiments approach associated with neo-institutionalism.<sup>56</sup> Historical analysis is conducted in reverse, on the basis of what is observed today. Institutional continuity, which in most cases is neither demonstrated nor critically examined, instead serves as a hypothesis for interpreting statistical correlations as the economic effects of institutions over the long term. This is an even greater problem when it comes to defining institutions and quantifying their characteristics, as must be done if the causal reasoning specific to natural experiments is to be pursued.

54. In public policy, too, it can be ill-advised and potentially dangerous to predicate policy implementation on the ability to assess it using statistical methods; see Stephen T. Ziliak and Deirdre N. McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives* (Ann Arbor: University of Michigan Press, 2008).

55. Naomi Lamoreaux, "The Future of Economic History Must Be Interdisciplinary," *Journal of Economic History* 75, no. 4 (2015): 1251–57, here p. 1255. Lamoreaux refers explicitly to Nunn's studies of African history cited in n. 64 below.

56. This criticism could already have been made of attempts by economists—including the first generation of new economic historians—to use statistical and theoretical tools to study history. However, Lamoreaux is right that, over time, some historian economists have either surrendered or adjusted their neoclassical assumptions and managed to tailor their quantitative procedures to more context-sensitive approaches; see Clément Dherbécourt and Éric Monnet, "Les angles morts de *The Poverty of Clio*," *Tracés* 16 (2016): 137–50.

## The Question of Measurement

The practice of quantitative history requires researchers to reflect on whether measurements are suited to the object being measured and to consider any statistical or cultural assumptions the figures may contain.<sup>57</sup> The natural experiments approach also runs into other problems, for two reasons. First, it compares statistics produced at different periods and therefore not really commensurable (trade and population data, for example). Second, and even more importantly, the desire to generalize leads researchers to make intensive use of proxy variables. For example, the ability to grow potatoes in a given type of soil becomes a proxy for potato-growing which in turn becomes a proxy for changes in nutrition, and it is this last factor whose causal effect on growth is then assessed.<sup>58</sup> In most cases, the validity of the proxies cannot be verified using other sources. The dependence on proxies raises the obvious risk of measurement errors, the extent of which is hard to evaluate. The approach also runs into the problem of category construction.<sup>59</sup> In one series of studies, a group of economists used legal systems (civil law versus common law) as proxies (or, in certain cases, as instrumental variables, which amount to the same thing when it comes to estimating “reduced forms”) for debt repayment and therefore for the law’s ability to contribute to the efficient functioning of a market economy.<sup>60</sup> In fact, historians have shown that the practices associated with different legal systems have diverged sharply over time and even within systems, a finding that invalidates the economists’ underlying research hypotheses and therefore their causal reasoning.<sup>61</sup>

It is important to emphasize that what is at issue here is not just a question of accurate or inaccurate measurement or of the statistical problems caused by measuring errors. Using proxies is actually a means of attaining a higher level of generality and so of affirming the existence of a causal relation when the

57. Witold Kula, “Histoire et économie. La longue durée,” *Annales ESC* 15, no. 2 (1960): 294–313; Jean-Yves Grenier and Bernard Lepetit, “L’expérience historique. À propos de C.-E. Labrousse,” *Annales ESC* 44, no. 6 (1989): 1337–60; Claire Lemerrier, “A History Without the Social Sciences?” *Annales HSS (English Edition)* 70, no. 2 (2015): 271–83.

58. Nathan Nunn and Nancy Qian, “The Potato’s Contribution to Population and Urbanization: Evidence from an Historical Experiment,” *Quarterly Journal of Economics* 126, no. 2 (2011): 593–650.

59. See Hopkins’ critique of using population data as a proxy for economic growth in Africa in “The New Economic History.”

60. Rafael La Porta et al., “Law and Finance,” *Journal of Political Economy* 106, no. 6 (1998): 1113–55.

61. Claire Lemerrier, “Napoléon contre la croissance ? À propos de droit, d’économie et d’histoire,” *La vie des idées*, 2008, <http://www.laviedesidees.fr/Napoleon-contre-la-croissance.html>; Jérôme Sgard, “Do Legal Origins Matter? The Case of Bankruptcy Laws in Europe, 1808–1914,” *European Review of Economic History* 10, no. 3 (2006): 389–419; Aldo Musacchio and John D. Turner, “Does the Law and Finance Hypothesis Pass the Test of History?” *Business History* 55, no. 4 (2013): 524–42.

econometrics show only a correlation, all else being equal. No one would dare claim, for example, that the French army's post-Revolution invasion of German territories resulted in greater growth due to France's superior economy. Yet that conclusion would be consistent with the econometric analysis conducted in the article discussed above. Only by considering the French invasion as a proxy for institutional change of a kind assumed to foster economic growth could the authors draw a causal interpretation from their correlation. Likewise, in African economic history, several studies employing natural experiments have relied on a measure of the social hierarchy within a given ethnic group as a proxy for the degree of political centralization (unaffected by colonization) in order to measure long-term institutional causal effects. But as Morten Jerven, Denis Cogneau, and Yannick Dupraz have pointed out, the ethnic categories used as proxies in these studies are likely to tell us more about past assumptions by the English anthropologists who constructed them than the political organization of African states.<sup>62</sup> The epistemological problems involved in using proxies are not specific to economic history, but they do arise with particular insistence in historical analyses based on an interventionist notion of causality. In such analyses, the "intervention" under study, inevitably removed from its historical framework, has to be defined and described in such a way as to allow researchers to interpret a statistical correlation in terms of causality. The link between proxy and causality is therefore inherent to this method of causal analysis.

### Temporality and the Compression of History

In a text that strongly resonated with economic historians of Africa, Gareth Austin denounced the "compression" of history by authors using the natural experiments method to carry out historical studies.<sup>63</sup> Having found a negative correlation between the magnitude of historical trade in slaves and current economic development, Nathan Nunn concluded that the first process had caused the second.<sup>64</sup> He hypothesized that this causal relation resulted from the political dissolution brought about by the slave trade. For Austin—in addition to his explicit skepticism about the reliability of Nunn's data—this hypothesis ignored the difference between the precolonial and colonial periods and was not founded on any tangible elements that might support linking the slave trade to African political development. Ewout Frankema and Marlous van Waijenburg later reformulated the same critique from a statistical perspective, pointing out that

62. See the thematic dossier "The Economics of Contemporary Africa," in *Annales HSS (English Edition)* 71, no. 4 (2016): 503–79; Jerven, "A Clash of Disciplines"; Denis Cogneau and Yannick Dupraz, "Institutions historiques et développement économique en Afrique," *Histoire et mesure* 30, no. 1 (2015): 103–34.

63. Austin, "The 'Reversal of Fortune.'"

64. Nathan Nunn, "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics* 123, no. 1 (2008): 139–76. See also Nathan Nunn and Leonard Wantchekon, "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review* 101, no. 7 (2011): 3221–52.

the negative correlation Nunn observed does not appear when pre-1970s economic development data is used.<sup>65</sup>

Central to these criticisms are questions of the temporality of the processes studied, the interaction and multiplicity of causes over time, and the analytical categories used. The problems in natural experiments are compounded by the fact that, in keeping with the canons of “applied economics,” their econometric models are linear and a-theoretical (nonstructural). This means that, even if we ignore the problem of category construction and accept neo-institutionalist presuppositions, the authors of these studies still have not taken into account non-linear interactions between the different factors contributing to a phenomenon, to say nothing of the possibility that those interactions may change over time. The linear model can be a good approximation of reality over a short period of time and at the microeconomic level when there are no major discontinuities in relations between the observed factors. But that is often not the case of long-term dynamics involving macroeconomic variables with complex interactions that may be subject to sudden change. The problem is made worse by the fact that these studies refer to institutions without observing institutional practices. The non-linearity of the relation between the scale of the slave trade in the past and the current gross domestic product of a country is a good illustration of these difficulties (and the types of non-linearity at work can be quite different from the one described here).

This is a familiar problem in macroeconomics, and econometric developments in this realm are vastly different from the “experimental” techniques imported from microeconomics into economic history. It is therefore paradoxical that the proponents of natural experiments in history tend to be drawn to subjects that are macroeconomic in nature, both in terms of their object (growth, trade, human capital and productivity, etc.) and their time-scale (the long term). Two macroeconomists—Edward Leamer and Christopher Sims—have reacted strongly to Angrist and Pischke’s claim that historical studies by Acemoglu, Robinson, and Nunn successfully introduced the natural experiments approach into the field of macroeconomics. Leamer reaffirmed that macroeconomic research could go no further than identifying patterns and linking them to theories: “I think that Angrist and Pischke are way too optimistic about the prospects for an experimental approach to macroeconomics. Our understanding of causal effect in macroeconomics is virtually nil, and will remain so. Don’t we know that?”<sup>66</sup> Sims, who has made important contributions to econometrics since the 1980s, reacted even more forcefully, stating that natural experiments were nothing more than “rhetorical devices that are often invoked to avoid having to confront real

65. Ewout Frankema and Marlous van Waijenburg, “Structural Impediments to African Growth? New Evidence from Real Wages in British Africa, 1880–1965” (Centre for Global Economic History working paper no. 24, Utrecht, 2011).

66. Edward E. Leamer, “Tantalus on the Road to Asymptopia,” *Journal of Economic Perspectives* 24, no. 2 (2010): 31–46, here p. 44.

econometric difficulties.”<sup>67</sup> He refers to two difficulties in particular, namely identifying the estimated econometric equation and correcting errors of estimation—which are technical translations of the limitations described above.

Historians’ critical remarks on the compression of history dovetail with those of macroeconomists on one fundamental point: with an interventionist definition of causality, it is not possible to affirm that an event in the distant past was the cause of a contemporary event if the complete chain of relations between the two has not been clearly identified. To fully refute the historical compression critique without surrendering the notion of causality, natural experiment economists would have to integrate a processual approach to causality in history, whereas, as we have seen, the natural experiment approach favors interventionist-type causality. In the philosophy of causality, Wesley Salmon is the most eminent advocate of the process theory of causality, which maintains that causality is a characteristic of continuous processes rather than a relation between events.<sup>68</sup> Salmon’s theory implies both the production and the propagation of causal influence, with causal influence being generated by way of causal processes. A causal process is defined as something that presents a coherent and constant structure over time.<sup>69</sup> If that coherence is not established, then it is inaccurate to speak of causality, regardless of the research design or techniques used.

Interestingly, process theories of reality are directly compatible with the idea of singular (that is, non-generalizable) causality—precisely the type defended by Weber in the social sciences. But they do not exclude causal pluralism (multiple causes), either. A causal process may consist in the manifestation of a singular relation between one or several causes and one or several effects. The idea that a causal process displays a form of structural invariance over time should not be taken to mean that it can be reproduced across history and therefore be generalized; rather, that invariance is a condition of its realization, which can be both singular and complex. Adapting a process theory of this kind might help proponents of the natural experiments approach to move beyond a notion of

67. Christopher A. Sims, “But Economics Is Not an Experimental Science,” *Journal of Economic Perspectives* 24, no. 2 (2010): 59–68, here p. 59.

68. It is beyond the scope of this article to present all the alternative conceptions of causality that have been developed in macroeconomics, but we can cite three—while noting that none is without problems or better than the others. Granger’s probabilistic causality is based on much weaker postulates than interventionist-type causality, stating simply that A causes B when A chronologically precedes B and when variations of A allow us to (partially) predict variations of B. Another procedure often used in macroeconomics is to estimate a stylized model of the economy and simulate exogenous shocks in that model in order to discuss how those shocks may correspond to observed variations. This type of causality is both interventionist and structural in that all interactions resulting from an exogenous shock are modeled. The third procedure, closer to process theory, is to document common statistical trends, or regularities, and study them in relation to general historical developments.

69. Wesley C. Salmon, *Causality and Explanation* (Oxford: Oxford University Press, 1998); Phil Dowe, “Process Causality and Asymmetry,” *Erkenntnis* 37, no. 2 (1992): 179–96.



causality that is excessively remote from historians' understanding of historical regularities or irregularities. Moreover, the process conception of causality is not, in principle, incompatible with statistical modeling of causality, although this avenue has yet to be explored.<sup>70</sup>

### Experimentalist "Nature"

Natural experiments are, by definition, rare: the real world does not resemble a laboratory experiment. This issue—the question of modality, in philosophy—cannot be elided in discussing the relationship of theory to reality. In other words, it is not simply a matter of observing that certain historical events can be described as natural experiments in reference to interventions in a laboratory setting; we must also ask why their occurrence is possible at all given that it is, by nature, abnormal. In this question, the philosophy of history and the philosophy of causality overlap.

If we consider a theoretical model to be a set of relations between variables—which can also be equated with a “thought experiment”<sup>71</sup>—what, then, is the connection between a thought experiment and a natural experiment? Is causality in the natural experiment the same as in the model? To answer this question, it is necessary to distinguish between two cases: either the exogenous variable observed in the natural experiment corresponds to an exogenous variable in the model, or it corresponds to what in the model is normally an endogenous variable. The first of these situations is often used to test economic models, particularly in microeconomics, where models usually have a relatively high number of exogenous variables.<sup>72</sup> Here, the exogenous variation has the same status in the experiment as in the model. In the second situation, the exogenous variable corresponds to an equilibrium variable in the model (the price determined by supply and demand, for example); that is, to a theoretically endogenous variable. This is often found in macroeconomics, whose models (called general equilibrium models) usually have endogenous variables only. Intuitively, we can sense the problem (of “identification,” as it is called) that arises when researchers studying a natural experiment observe an exogenous variation of a variable, while they are in fact referencing a model in which the variable is endogenous: the natural experiment has been taken for a real equivalent of the thought experiment in that,

70. Phil Dowe, “On the Reduction of Process Causality to Statistical Relations,” *British Journal for the Philosophy of Science* 44, no. 2 (1993): 325–27.

71. On the analogy between models and thought experiments, see Mary S. Morgan, *The World in the Model: How Economists Work and Think* (Cambridge: Cambridge University Press, 2012). This analogy is often used by economists themselves; see Leamer, “Tantalus on the Road to Asymptopia,” 44.

72. This is because microeconomic models are models of partial rather than general equilibrium. See Steven D. Levitt's test of a standard economic model of crime, which uses an exogenous change in the probability of the crime being identified: “Testing the Economic Model of Crime: The National Hockey League's Two-Referee Experiment,” *Contributions to Economic Analysis and Policy* 1, no. 1 (2002): 1–21.

normally, only the mind (or digital simulation) can “exogenize” a model’s endogenous variable. This problem has a long history. Even as David Hume was devising the first economic-model thought experiment in the form of a counterfactual statement, he had already imagined the problem in his 1752 text, *Of the Balance of Trade*. Indeed, he concluded that there was a logical contradiction in the very idea of a natural experiment reproducing a thought experiment: “The same causes [that would create the natural experiment] ... must prevent their happening in the common course of nature.”<sup>73</sup> In other words, because the model supposes that all variables are endogenous to each other, the occurrence (or hypothesis) of an exogenous relation is contradictory to the model. We therefore have to either change the model or ensure that the model itself offers a theoretical justification for the exogeneity of a variable-to-variable relation.<sup>74</sup> Pearl’s causal inference theory involves such a hypothetical move: to turn a variable that a theoretical model assumes is endogenous into an exogenous variable.<sup>75</sup>

Pearl’s understanding of causality is also interventionist and is fairly close to the one used in the natural experiments method, but it gives us a clearer view of the link between exogenous intervention on the one hand and modality and

73. Hume, “Of the Balance of Trade,” *Political Discourses* II, here 5.11. Hume argued that the balance of payments would always achieve equilibrium due to the link between money supply, prices, and trade. After constructing a model interlinking these variables, he devised the following thought experiment: “Suppose four-fifths of all the money in GREAT BRITAIN to be annihilated in one night, and the nation reduced to the same condition, with regard to specie, as in the reigns of the HARRYS and EDWARDS, what would be the consequence?” (*Political Discourses* II, 5.9). Once he had used this thought experiment to isolate and describe the theoretical mechanisms at work, he clarified that what he imagined could not happen in reality: “Now, it is evident, that the same causes, which would correct these exorbitant inequalities, were they to happen miraculously, must prevent their happening in the common course of nature, and must for ever, in all neighbouring nations, preserve money nearly proportionable to the art and industry of each nation” (*Political Discourses* II, 5.11).

74. In “Chronicle of a Deflation Unforetold,” *Journal of Political Economy* 117, no. 4 (2009): 591–634, François R. Velde studied the effect of an arbitrary reduction of the money supply in France in 1724, an event that might seem to correspond to Hume’s thought experiment. But he refused to identify this as a natural experiment because, in eighteenth-century France, it was customary for the king to make this type of decision and economic actors took that into account.

75. Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge: Cambridge University Press, 2009). Pearl’s conception of causality as it relates to econometrics is based on older works by Trygve Haavelmo. Specifically, the aim is to distinguish clearly between a condition and a cause, the latter involving fixing a parameter in the model (and therefore exogenizing it from the set of equations). By contrast, a condition is defined as a purely statistical concept. Pearl’s approach is a reminder that what statisticians and econometricists estimate is only ever a statistical relation and therefore a condition. To have causality, researchers must make an additional hypothesis; that is, they must explain why the variable can be thought of as “fixed.” For Pearl’s own critique of the confusion economists perpetuate on the statistical handling of causality, see Bryant Chen and Judea Pearl, “Regression and Causation: A Critical Examination of Six Econometrics Textbooks,” *Real-World Economics Review* 65 (2013): 2–20.

counterfactuality on the other. Pearl takes the concept of (possibly impersonal) causal intervention as a primitive analytic proposition enabling him to define an exogenous causal event or process (exogenous to that which it is assumed to influence or change). This approach also places the notion of exogeneity at the center of his definition of causality: later manipulation of the effect does not retrospectively affect the event or process corresponding to the cause. We may, of course, doubt whether exogenous variables are possible in history, but that difficulty is not logically insurmountable, and it is not the thorniest problem here. It is possible to correlate different causes, even if they stand in a hierarchical relation to each other; in that case, the exogenous variable simply becomes complex. The delicate part, rather, is the very idea of exogenizing a variable on the basis of a model of historical reality. And as we saw above, without a model, it is harder to generalize and interpret historical data. In Pearl's understanding, the idea of causality, and therefore of an exogenous variable, is analytically associated with the idea of counterfactuality. If  $X$  causes  $Y$  (univocally in this case, let us say), then if  $X$  had not been the case,  $Y$  would not have been the case either. Using this counterfactual formula enables us to grasp one of the fundamental aspects of our intuition about causality (along with the postulate that cause precedes effect). Yet that intuition seems to contradict what we think of as a historical event. For it means that an exogenous historical event  $X$  is such that, if it had not taken place, its effects would not have been the case; yet  $X$  is the case and its effects are real. So the counterfactual turns out to be what would have happened if history had followed its course—which it did, necessarily. The apparent paradox, then, is that, independently of any intervention from an exogenous variable that changes it, the course of history has become the counterfactual reference point of an experiment that takes place in reality. Proponents of the natural experiments approach resolve this apparent paradox by delimiting intervention and non-intervention zones within geographic and historical reality, on the model of a laboratory experiment. Clearly, though, the idea of an exogenous causal variable, associated as it is with a counterfactual notion of causality, suffers conceptually from this difficulty, which is a matter of modal semantics, for it leads to accepting different planes of reality *within* history.

When the model is perceived as a “thought experiment,” then, the relations between causality and history and between causality and economic theory remain indeterminate. Consequently, natural experiments do not allow us to resolve the paradox Hume observed. That the method may be consistent and logically justified, as Pearl's reasoning suggests, does not save it from the problems of identification that make it impossible to use this type of experiment to test a precise, specific theory. Hume's reasoning involves questioning why an event happens exogenously if it is defined as endogenous in the theory. Is the validity of our available models called into question by the simple fact that they do not acknowledge or explain the possibility of such exogeneity? This question should be addressed to the institutionalist theories that form the basis of so many of the interpretations in the articles presented above.

Studying Pearl's model would enable proponents of natural experiments in history to bring this difficulty to light formally and to imagine possible conceptual solutions to it that would move beyond simply attempting to prove the exogeneity of the cause. But it could also lead them to detail their implicit conception of causality and to distance themselves from the interventionist paradigm or discard it altogether. For example, is it necessary, in order for an event within a model to be considered as a cause, for it to be entirely exogenized regardless of the situation in which it is naturally positioned? The notion of "INUS causation" put forward by John Mackie in 1965 (prior to Pearl's work) might be adapted for the purposes of analyzing causality in history.<sup>76</sup> A cause is an INUS condition, meaning that it is an insufficient but necessary part of a condition that is itself non-necessary but sufficient with regard to certain identified effects. The fact that a condition can be a set of factors necessary to a condition that is itself non-necessary yet sufficient offers two ways out of the difficulty we have noted. A cause is not isolated; it is part of a complex set that we call the condition of an effect. It is determinant as a part of this condition, but the condition itself is not necessary for the effects to be realized. This considerably weakens the modal constraint—and ensuing difficulty—at the core of the interventionist conception. For example, it allows us to say that the drought that occurred in the summer of 1788 may be thought of as an INUS cause of the French Revolution in that the climatic conditions that year and the famine that followed are a necessary but not sufficient part of the economic explanation of the event. The famine is not solely explained by the drought but also by other factors: an ineffective farm resource distribution system, for example. The economic conditions linked to the climatic factors that prevailed on the eve of the Revolution are, in turn, a non-necessary yet sufficient condition (the French Revolution can be explained by other, non-economic factors). A cause is therefore one of a configuration of factors. Moreover, we can use resemblance and comparison to study such configurations in and of themselves. Charles Ragin identified three main approaches in the social sciences: research into individual cases, "configurational" and "comparative" research, and research oriented by the discovery of causal variables.<sup>77</sup> Clearly, when economists turn history into a field of natural experiments they are spontaneously adopting the third approach. Yet the second would surely lead them to a notion of causality more intuitively relevant for historical analysis.

Let us stress once again that the question of the counterfactuality and modality of an experiment is entirely different from the widely addressed question

76. John Leslie Mackie, "Causes and Conditions," *American Philosophical Quarterly* 2, no. 4 (1965): 245–64. INUS is the term used in the international literature to designate Mackie's definition of causality: "Insufficient but Non-redundant parts of a condition which is itself Unnecessary but Sufficient."

77. Benoît Rihoux and Charles C. Ragin, eds., *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (Thousand Oaks: Sage Publications, 2008). We are much beholden to one of the anonymous peer reviewers of this article for suggesting how Ragin's analyses could help imagine alternative solutions to the type of causality envisioned by proponents of natural experiments.

of whether an event was truly exogenous with regard to other variables. The point here is to understand why an event occurs and what its occurrence tells us about a given historical period and existing models. The counterfactuals on which natural experiments are based are often inaccurately formulated in that they do not articulate the question of the historical meaning of what is posited as an exogenous intervention. To return to the example of the study concerning the French Revolution and the Prussian territories, we have seen that the authors seem to have hesitated between the following counterfactuals: “Without the French invasion, the invaded territories would not have experienced greater growth than non-invaded ones” and “If the French Revolution had not taken place, the institutions of the ancien régime would have remained identical in the different Prussian territories.” What the advent of the French Revolution itself tells us about ancien régime institutions in Europe was not perceived as important for measuring its effects. And yet the very occurrence and existence of the Revolution shows that ancien régime institutions could be criticized and even overturned. Taking this into account means thinking about the various possibilities of change within those institutions (in France and elsewhere), independently of the exogenous changes sparked by the French conquests. And doing so calls into question the study’s very assumption that an external institutional shock was necessary to economic development.<sup>78</sup>

## Taking Account of Multiple Causalities

Centuries of philosophy and science have not produced a unique, consensual definition of causality, and it is unlikely that the definition used today by a particular group of economists is the finest one there is.<sup>79</sup> Interdisciplinary dialogue would have much to gain from a general recognition of that plurality (which, for economists, is foremost a function of the analytic scale and explanatory paradigm they have chosen), to say nothing of further conversation on causality and how to use it in historical study, in order to prevent any one definition from gaining a monopoly. Do the research and debates mentioned in this article attest merely to intra-economics disputes and divergences that historians can just as well ignore? We do not think so, for at least three reasons.

First, the claim to scientificity of authors seeking to apply to history the methods of natural experiments and of statistical analysis of long-term effects has been conceived *in opposition to* “the traditional methods of historians,” which are thus considered less rigorous and ambitious. This direct confrontation has

78. In his review of Diamond and Robinson, *Natural Experiments of History* (p. 274), Mokyr recalls that it was possible to adopt “progressive” institutions even where there was no French conquest and that Spain instated no reforms despite the French invasion.

79. In “Causation in the Social Sciences: Evidence, Inference, and Purpose,” *Philosophy of the Social Sciences* 39, no. 1 (2009): 20–40, Julian Reiss explains that the natural experiment method is based on an interventionist conception of causality.

already sparked debate—at least in the United States—about what economic history should be today, and it is not hard to imagine that it will impact where academics publish and the balances of power between disciplines within universities.<sup>80</sup>

Second, franker and more constructive interactions might gradually develop between the type of studies presented in this article and different historical approaches, along the lines of what happened with earlier neo-institutionalist research by North and Greif. This is all the more likely when studies based on natural experiments rest on extensive archival work that brings to light new data, or even new hypotheses, that can in turn be of use to historians.<sup>81</sup> The question of causality, specifically as it relates to temporality and the long term, together with the possibility of contextual causality, would move to the center of discussions between the two disciplines, as was the case more than a century ago and as occurred in recent debates on how the notions of economic rationality and the market might be used in historical analysis. As Claire Lemerrier suggested in a recent issue of the *Annales*, when researchers avail themselves of a long-term approach—which some say is on the verge of a revival—they have no choice but to reflect on the matter of causality.<sup>82</sup> We have tried in this article to show how such debates might be informed by philosophical work. There are alternatives to an interventionist, non-processual, decontextualized notion of causality. The criticisms formulated here of certain ways of using causality in history should not be read as a call to ban causal analysis from historical analysis but rather as a plea to reflect on the possibility of taking historical temporality into account without abandoning causal reasoning.

Last, we have alluded several times to the strong political implications of studies on the causal effects of institutions over the long term. Understanding the rhetoric and arguments used in those studies, as well as their implicit conceptions of the causality of history and institutions, will equip us to develop counter-arguments. The same can be said of the role of historical analysis in the social sciences, and in the polis at large.

*Sacha Bourgeois-Gironde*

*Université Paris 2 and Institut Jean Nicod*

*Éric Monnet*

*Banque de France, École d'économie de Paris, CEPR*

80. See the references listed in n. 1 above.

81. This is probably the case for some studies on the effects of Protestantism.

82. Lemerrier, “A History Without the Social Sciences?”