Complexity And Causation

Timothy Burke
Swarthmore College, tburke1@swarthmore.edu

Follow this and additional works at: https://works.swarthmore.edu/fac-history

Part of the History Commons
Let us know how access to these works benefits you

Recommended Citation
https://works.swarthmore.edu/fac-history/2

This work is brought to you for free by Swarthmore College Libraries' Works. It has been accepted for inclusion in History Faculty Works by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.
THE RELATIONSHIP BETWEEN cause and effect is one of the central objects of investigation in most modern academic disciplines. Disciplines which are especially sensitive to or interested in the passage of time are particularly concerned with causation, and of those, perhaps history most of all.

This is not to say that historical scholarship must necessarily be focused on arguments about causation and causality. Many works of history are more descriptive and aim to provide less an account of the why of historical change than an account of the what of history. In a more elaborate vein, historical writing influenced by postmodernism, particularly by the work of Michel Foucault, remains wary of causal argument and its tendency to look for origins, preferring instead what Foucault called “genealogy,” a style of historical writing focused on process and development over time without recourse to causal arguments which situate themselves outside of or underneath the history being described.¹

Historians face some distinctive problems in dealing with causation. With the exception of economic or other quantitative historical research that is dealing with extremely rich and rigorously collective data, most fields of historical study must make causal arguments without the ability to repeat experiments and without statistical tools like regression analysis that allow other disciplines to select among a host of competing variables to isolate and describe the relative magnitude of various causes of observed effects. Causation for historians is largely a matter of persuasive argument. As R.G. Collingwood put it, historians do not identify events and then ponder their causes separately, as scientists do. When a historian “knows what happened, he already knows why

Tim Burke is Associate Professor of History at Swarthmore College.

Soundings 90.1-2 (Spring/Summer 2007). ISSN 0038-1861.
it happened” because cause, event, and effect are unveiled through the work of interpretation rather than empirical discovery. To some extent, such arguments among historians can resemble or draw usefully upon the more rigorous approach to causality in analytic philosophy, but clear definitions of events, effects, and causes themselves in history necessarily refer to profoundly contestable, fuzzy, or ambiguous phenomena. As the philosopher Maurice Mandelbaum has observed, arguments about causation in formal historical scholarship often tend to draw upon a diverse mix of philosophical and everyday antecedents, often implicitly so.

Broadly speaking, causal arguments among historians since the mid-twentieth century have tended to divide into two approximate camps. The first set of arguments operate at long scales of time, refer to underlying structures or forces, and are relatively deterministic in nature. The second tend towards short scales of time, refer to specific and granular events or episodes, and often stress the relatively contingent or unintended nature of causality. There are sophisticated strategies for integrating these levels of analysis, such as Anthony Giddens’s theory of “structuration” that describes agency and structure as a dynamically recursive relation in which the contingent or unpredictable actions affect the more fixed or deterministic character of social structure, and vice-versa, in an endless feedback loop. Even with such integrations (not dissimilar to efforts by “compatibilist” philosophers to integrate free will and causal determinism), many individual historians still tend to prefer operating at one or the other scale of causal explanation.

In both cases, arguments about causation tend to run into serious problems. For the historian who emphasizes the vast expanses of time of the longue durée and the determinate role of underlying structures, there is a fundamental question, posed best by Salman Rushdie in his novel The Satanic Verses: How does newness enter the world? If change over time is best understood as determined by underlying and highly deterministic structures, and best perceived in long scales, why should there ever be anything new in history? Why should there have been capitalism, or the expansion of Western Europe, or the French Revolution?

There are a variety of ways to approach this basic problem. One is to deny or deemphasize the extent to which anything is,
in fact, new. For example, the social scientist Andre Gunder Frank has argued that the seeming break in global economic and political history that coincides with the rise of Western European societies to worldwide domination after 1500 and the spread of capitalism is largely an illusion, that there is only one global system which is 5,000 years old.\(^6\) Most specialized fields of historical scholarship feature work with related strategies to de-emphasize or even wholly erase a perceived break or disjunction between eras or systems, or which argue against the understood novelty of particular events, a mode of historical writing which is sometimes described as “revisionism.”\(^7\) For example, the field of medieval European history in recent years has engaged in a debate about whether there was ever such a thing as “feudalism.”\(^7\)

Another argument is to reduce strongly the moments of genuine newness or novelty in global history to a small handful of important eras, or possibly even to a single instance, in a perceived divide between the modern and premodern eras of human history. Perhaps one of the most fertile and intensely debated bodies of historical knowledge is concerned with the underlying causes of modernity. The fewer the cases of actual “newness,” the less difficulty they pose for strongly deterministic, large-scale accounts of causality, and the more that such novelty can be represented as a kind of singularity, an unusual instance of a large-scale contingent outcome arising from underlying social structures. Versions of this strategy would include strongly teleological interpretations of history, most notably Marx’s historical materialism. In this view, while there may be both new eras or periods in the history of particular societies and even short-term events which are novel or disjunctive, such newness arises from deterministic forces which guide overall historical change towards a predictable end. Relatedly, even non-Marxist historians who emphasize the long-term determinate role of materialism sometimes argue that short-term and contingent technological, environmental, or biological events give rise to social, cultural, economic, or political novelty within human societies. An earthquake which strikes a large city at a particular moment in its long-term evolution may produce far-reaching consequences which would not have necessarily followed had a similar earthquake struck the same city at a different historical moment.
Here more deterministic or large-scale arguments about causation may end up casting a light on some of the problems that short-term and contingent approaches to causation in history face. The historian who presents a fairly deterministic argument about the history of a given technology often ends up obscuring the particular contingencies of its creation. A technology may be discovered, refined, and propagated through a series of relatively chance events. This technology may then lead to variant outcomes depending on where and when it comes into being. Seen from the *longue durée*, it may appear inevitable that early modern reorganization of textile production in northern Europe would lead into the industrial revolution in England which would lead to the industrial production of guns which would lead to the early machine gun which would give European armies a decisive military advantage in early twentieth-century Africa which would lead to the partition of the continent under European control, and so on. But seeing all this change over time as the causal expression of long-term deterministic forces tends to obscure the long chance of seemingly unintended and often distinctively local or specific histories along the way. For each one of these histories, different outcomes seemed quite plausible. Break any link in such a long and elaborate chain, and serious problems of counterfactual reasoning tend to spring up.

On the flip side, however, histories that dwell on events, on local scales of human interaction, and on the deliberate actions of human agents, tend to go begging for true causal arguments and often rely upon mere precedence. They simply assume, that is, that the causal explanation for a given event is to be found in the event which immediately preceded it. At this scale, explaining newness is no trouble at all: Human agents seem highly capable of invention or imagination; institutions and everyday practice are demonstrably plastic; and events readily give rise to unpredicted or unintended consequences. The historian’s problem here is to explain continuity and repetition, to explain why events should consistently turn in a particular direction across a large spatial or temporal scale, or why similar patterns of historical change should appear in disparate locales or cases.

In both cases, moreover, historians who wish to talk about causation face a basic problem with the rhetorical form of historical knowledge, which is almost invariably built around persuasive hu-
manistic writing. In such a format, crafting genuinely multivariable causal arguments is profoundly difficult. While most historians recognize the intellectual danger posed by what Marc Bloch called “the fetish of the single cause,” in practice historical argument at both the long and short scales of focus tends to make causal claims in terms of a very small handful of discrete events or underlying structures. Newly minted doctorates may excel at entering into long-running debates and pronouncing them to be “more complex” than previously appreciated, but there are limits to this kind of gesture. A work of historical scholarship which pronounced a given event or effect to be irresolvably and infinitely multiple in its causation would add little to what we know, unless it were a general philosophical assault on all discussions of causality in historical study. At the same time, most historians recognize that emphasizing one or several causes (at whatever scale or level of determination) is ad argendo, a necessary exaggeration or abstraction to permit the discussion to go forward.

Emergence and Historical Causation

I believe that the phenomenon known as “emergence” and a body of associated concepts such as complexity, complex adaptive systems, networks, and agent-based systems hold considerable promise for thinking about causation in historical study. These concepts cannot resolve any of the problems that I have described thus far. These difficulties are intrinsic to historical reason and will always remain contentious. Emergence, however, casts some of these problems in a new light, or allows an approach from some new angles. In the long term, it may also offer new forms of historical representation, new rhetorical or argumentative instruments. At the same time, emergence in historical thought runs into exactly the same epistemological and practical issues which sharply limit its potential usefulness in most, perhaps all, fields of human knowledge.

The simplest definition of emergence as a phenomenon is that it is a process of change over time in which complex systems, patterns, or structures form in an unplanned or undesigned manner from simple or disorganized initial conditions. Typically, emergence results from the autonomous and simultaneous interaction of a very large number of independent agents, each pos-
sensing a set of rules determining its actions within a particular environment.

The free software NetLogo is a good platform for simulating emergence as a process. Let me describe one NetLogo simulation called “Termites.” In it, there are two differently colored dots, each one pixel on a computer screen. One dot is a “termite,” the other a “stick.” Both are randomly distributed at the start of a simulation within an otherwise empty environment. The sticks are merely environmental: They do not move or act. The termites act, governed by a set of simple rules. Each step of the simulation, they move one space in a random direction. If they end their move proximate to a stick, they “pick it up” (indicated by the termite changing color). If while carrying a stick, they end their move next to yet another stick, they drop the stick. That’s more or less it. In any given simulation of “Termites,” with a random distribution of termites and sticks, the termites will eventually build a single round “pile” of sticks. (The display in NetLogo wraps around, so what may appear to be two “piles” at the top and bottom of the screen are in fact a single one.) This pile is a permanent feature of the environment once it appears: It will never be pulled apart. No other pile or structure will appear. Its shape, once formed, is relatively stable. But the termites don’t have any instruction to create a pile. There is no master agent governing their actions. There is no concept of a pile in their rules or in the environment.

![Fig.1 Three stages of “Termites” in NetLogo](image)

Some tangible or empirical examples of the concept of emergence or self-organizing systems include the movement of social insects around obstacles, patterns of coordinated growth among slime molds, the coordinated action of bird flocks in flight, and
the formation of cloud patterns within Jupiter’s atmosphere. More contentiously, some scientists have argued that human consciousness, evolution, and the large-scale structure of matter in the universe are all examples of emergent systems, where complex behavior or structures at one level have arisen from the interactions of simple rules determining the behavior of many autonomous agents or components at a lower level of organization.

What should be clear from the outset, however, is that using the concept of emergence in historical argument about causality potentially offends (or possibly complements) both strongly deterministic and strongly contingent inclinations in historical writing. It may be easiest to see what this means through specific examples and in so doing, begin to sketch which kinds of historical phenomenon are most richly served through the use of the concept. Let me start with a simple example: the technological history of the videotape player, most particularly, the eventual triumph of the VHS format over its rival Betamax format. This is an example which has drawn considerable interest from economic historians, historians of technology, and economists because most observers consider the Beta format to have been technologically superior in many respects to VHS. As one group of scholars describes it, this story has some familiar elements to it in the context of business history: the relation between “first movers” and later entrants to the market, the influence of marketing strategies, and the struggle to establish technological standards. Given that both formats had the backing of sizeable corporations with considerable power to influence consumer behavior, this cannot be seen as a “pure” case of emergence (and indeed, in human history, I would argue that no event or phenomenon could ever be so). But the end result in many ways seems unpredictable from the initial entry of Beta format into the marketplace. This result was produced by the simultaneous activities of many autonomous agents ranging from marketers to consumers. The crystallization of preference for the VHS gave rise to a new kind of complexity in media consumption practices and ownership which no institution or interest controlled or anticipated at the outset of the introduction of the technology. In general, the concept of emergence seems to be particularly helpful in understanding the history of technology. For example, it
has obvious applicability to the history of the Internet, and to the spread of cellphone usage throughout much of the developing world in the late 1990's and early 2000's.\textsuperscript{10}

The concept might also be productively used to compare the formation of similar or related industrial or economic systems. For instance, in southern Africa, the social history of gold mining since the 1870's has been the subject of considerable historical research; copper mining in northern Zambia and the southern Congo since the 1930's is less studied, but still relatively well-known.\textsuperscript{11} The differences between the social formations and labor systems in the two cases are substantial. Gold mining in South Africa was an important root of racial segregation and apartheid. The system treated workers as migrants who lived under tight control in guarded compounds during the term of their labor contract, and they were compelled to return to distant rural home areas when they were not working in the mines. Copper mining in northern Zambia and southern Congo, in contrast, relied upon "stabilized" labor, workers who lived with their families in permanent urban communities built by the mining companies around the sites of the mines themselves. It is possible to understand this difference in deterministic (often materialistic) terms. Gold mining in South Africa required lots of cheap but unskilled labor due in part to the nature of the gold deposits themselves and also, before the end of the gold standard in 1931, to the fixed price of gold in the world market. Copper mining required skilled labor, and the prices for copper were relatively favorable as African production was growing in size in the 1940's and 1950's. It is also possible to see the difference as a temporal one (that copper mine management had learned some lessons from the problems of the labor system on South African gold mines) or as an ideological one. (There were few white residents around the copper mines, but many in the Witwatersrand in South Africa.)

Thinking about both systems as emergent phenomena does not displace the explanatory value of these arguments, but it does add a useful element to the comparison. The compound system of South Africa's gold mines, for example, stems in part from the earlier development of diamond mining. Some of the controls on African workers in diamond compounds were governmental concessions to individual white prospectors who were
using the cheap labor provided by a system of apartheid as a hedge against very thin economic margins (as also happened in California when prospectors needed cheap migrant labor after the initial wave of the 1849 gold rush). Other controls, such as putting iron gloves on African miners when they ascended from the diamond mines, were crude attempts to deal with the problem of diamond smuggling, but quickly metamorphosed into more expansive systems of social control. Some aspects of the South African migrant labor system of 1900 could legitimately be said to be the compounded results of hasty improvisations and idiosyncratic initiatives during the initial development of industrial mining in that area. Equally, “stabilization” as a policy in the Copperbelt could be attributed to the intellectual and managerial backgrounds of key planners and executives involved in the initial development of the copper mines, and these backgrounds were, in turn, the product of industrial sociology in Europe and the United States during that era.

There are numerous examples of historical events, at varying levels of scale, that could be illuminated by a consideration of the role of small or incremental actions by many agents working independently of one another in creating larger and more complex systems that none of the actors intended to create. For instance, various treatments of “proto-industrialization” in Western Europe, such as Pat Hudson’s study of wool textile production in England in the eighteenth century, suggest that the transition to industrial capitalism was marked by a complex relation between quite different but successful systems of producing. In the Hudson case studies, there was a complex interaction between technologically and industrially oriented “worsted” industry and a more artisanal woolen industry, with the latter often surprisingly outpacing the productivity of the former. Hudson’s analysis seems to me to reveal the emergent character of the industrial system of textile production. The history of these two systems is in some sense the contingent and unplanned result of their interaction over time.

At an even higher level of scale and abstraction, I would argue that there are perennial theoretical questions in historical scholarship that benefit from a consideration of emergence and complexity. In my own recent work, I have been arguing for a reinterpretation of causal roots of the “new imperialism” of the
late nineteenth century, one which builds on both older and newer scholarly approaches to the issue. The “new imperialism” led to the seemingly abrupt division of almost all of Africa and parts of Asia and Oceania into the territorial possession of a handful of European powers. Historians have long puzzled over the reasons for this surge of imperial ambition, particularly given how relatively short-lived the resulting empires turned out to be. In my own work, I am less concerned with the overall phenomenon than the structure of the colonial governments put into place in Africa by England, France, Germany, and Belgium in the first decade of the twentieth century. These governments were typically constructed around principles that the British called “indirect rule,” a proposition that Africans would govern themselves but under tight constraints. They would have their own chiefs, but imperial authorities would select and control those chiefs. They would have their own customary laws, but those laws would be codified and selected by colonial officials. And so on.

Many scholars studying Africa have chosen, to varying degrees, to see indirect rule as a highly designed system intended in both its generalities and particulars to maximize the capacity of imperial rulers to dominate African societies. \(14\) An approach sensitive to emergent phenomena suggests instead that the system of indirect rule was the convergent result of many simultaneous and parallel improvisations and initiatives taken by both imperial and African agents in the chaotic period following the pronouncement of empire at the end of the nineteenth century. The resulting norms and practices of colonial administration were in this view unplanned to some significant degree, and their contradictions were an unintended consequence of the improvisational process by which these governmental practices were created. This view has become more common recently among historians of colonial Africa, such as David Gordon. \(15\)

This characterization of indirect rule runs into several problems. First is the history of imperial administration: Many of the bureaucratic procedures and processes put into place in African territories between 1870 and 1900 had precedents in other regions of the world such as India and the Caribbean or were based in some fashion on bureaucratic structures within Europe itself. Again, if emergence offers anything to the understanding of causality in historical scholarship, it is not as a “theory of every-
thing” which displaces all other frameworks or empirical knowledge. More importantly, if some of the bureaucratic and political processes characteristic of indirect rule in Africa were emergent phenomena, how did they become systematic, converging simultaneously out of many different localities into a single large structure of political power? Emergence theory as a way of thinking about causality really shines in answering such questions, because it is centrally concerned with processes in which organization emerges from relative chaos, in which complex structures arise from simpler and more inchoate practices. If one thinks about imperial administrators, African elites, peasants, white farmers, and other discrete groups of agents in early colonial Africa as having “rulesets” that shaped their actions (much as the “termites” in NetLogo), it is quite possible to see how the simultaneous interactions of their differing priorities could converge into a large-scale system without any of the actors necessarily intending to create that system (again, much as the “termites” in NetLogo create a circular pile).

There are many large-scale events in human history that might similarly benefit from incorporating notions of emergence into debates about the causal roots of those events. Events which are characterized by seemingly sharp and expansive discontinuities or transformations of social, economic and political life over a short span of time seem especially suited to this approach. The evolution of the French Revolution from the world of Parisian salons, rural discontent, and aristocratic decline in the late ancien régime of France to the tumult and uncertainty of the initial overthrow of the monarchy and then to the Terror has been endlessly analyzed by historians. Many have sought to relate the Revolution to deep and relatively deterministic structural causes and others arguing for the relatively novel (but deliberate) character of the Revolution. Surely at least some of the story is equally well-described as emergent, about the unintended consequences of divergent activities by many pre-revolutionary actors in France, with the excesses of the Terror being as much a surprise and puzzle to the participants (and yet completely comprehensible as an emergent consequence of the earlier history) as to any later observers.

Is emergence theory simply one more analytic perspective in the spectrum of approaches historians can employ? Perhaps, but even if not, it is a significant enough good in its own right. Emer-
gence is “good to think.” Even when it is not a sufficient explanation in and of itself, it can serve as a system for identifying causal explanations and interpretations which might not intuitively occur to a historian. Thomas Schelling’s 1971 mathematical modelling of residential segregation, now a staple in the scholarly literature on complexity and self-organizing systems, is a good example of this value. Schelling’s model suggests that segregation could potentially occur simply from the residential preferences of individual agents in a confined space, e.g., that it does not require top-down enforcement or organization. This doesn’t mean that this is in fact what has historically happened with residential segregation in the United States or elsewhere, but it does open up a somewhat counter-intuitive hypothesis to explore in the context of existing historical scholarship about race and spatial segregation. Emergence in this context is a kind of analytic black box to pass our causal assumptions and arguments through to see if some unfamiliar or non-intuitive explanation or idea presents itself on the other side.

Emergence theory also looks anew at the relationship between variables and outcomes in change over time. Most arguments about contingency in historical scholarship have to take the form of “for want of a nail,” a narrow kind of counterfactual reasoning or argument in which the historian imagines a single variable being different and then follows a chain of consequences flowing from that difference. Even with cases where different outcomes are highly plausible (say, for example, Lee winning at Gettysburg), the chain of counterfactual assertions becomes very difficult to follow as it progresses from a single difference to larger and larger scales of historical transformation. From Gettysburg, it is relatively easy to get to a victory for the South in the Civil War. But from a victory for the South, it is quite difficult to go any further. Could a slave system of agricultural production have survived the competitive force of Northern industrial capitalism? Or survived its own internal pressures? Could the South have maintained political cohesion? The emergence model, however, does not measure the relationship between initial conditions and structural outcomes through single chains of causality, but through the massive simultaneity of agents acting independently of one another within a constrained environment or space. I can imagine, if only as an experiment, a “counterfactual engine” or
agent-based simulation-machine which could allow a historian to model or describe divergent historical outcomes involving hundreds or thousands of variables in motion at once.\textsuperscript{17}

In this respect, the introduction of emergence to arguments about historical causality functions similarly to the intervention of Stephen Jay Gould into evolutionary theory in his book Wonderful Life, which contends that the evolutionary process is shaped a great deal by accident and contingency.\textsuperscript{18} The key thing about Gould's argument is that it relates contingency at the microscale of organisms and their composition to the macroscale of evolution and ecosystems as a whole. In history, scholars who place emphasis on contingency and agency often tend to do so against large-scale forms of determinism, asserting the autonomy of the individual and collective human subject and the importance of variable outcomes that derive from choice or willful action. Emergence does not reject that emphasis, but it puts that kind of contingency back into a new kind of relation to outcomes at larger scales.

However, this relation is also the conceptual Achilles heel of emergence and complexity theory. Emergence occurs when many agents and forces acting simultaneously within a constrained environment give rise to some new complex structure which then alters the environment within which those agents carry out their activities. When historians try to make causal arguments by focusing on a single variable, or a small handful of variables, most would concede that they argue reductively out of necessity. Most of us know that the world is more complicated than that, but it is very difficult to describe the causal relationship between two equally complex states with all the many variables that comprise them kept in view at once. Some kind of reductionism is an intellectual and rhetorical requirement. Emergence attractively envisions an asymmetrical relationship between one relatively simple state of affairs and a consequently more complex one: It gets to have its reductionist cake, but eat it too, to pose a dynamic causal relationship between simplicity and complexity, agency and structure.

An emergent view of indirect rule in Africa or the French Revolution can hold that simple, unintentional, and simultaneous interactions between a heterogeneous collection of human agents and institutions could give rise to novel large-scale politi-
cal and social systems. However, an emergence-based approach leaves us no way to understand the relationship between any particular agent or variable in the initial conditions and the resulting complex structures appearing at a later date, precisely because those structures come from the totality of all interactions between a large number of agents, institutions, and forces. The science writer John Horgan, writing skeptically about whether there are major new fields of scientific insight or knowledge which remain unknown, has argued that “chaoplexologists” (scholars studying emergence, chaos and complexity theory) have hard limits to the applicability of their theories.\textsuperscript{19} Certainly this seems the case with emergence in historical argument. Emergence allows for novel insights into the relationship between short-term contingency and long-term structure, but it also erects an epistemological veil between initial conditions and resulting complex structures. A historian can thus concede that it is entirely possible that any given action, agent or particular event was a “tipping point” that pushed one historical situation into a completely different systemic state, the way that water undergoes a phase change to ice. But a historian thinking about causality in emergent terms cannot know which agents or actions will produce which given systematic consequences because that is unknowable by definition.

For historians, that is less of a difficulty than it is in the hard social sciences, which tend to promise to measure the differential causal role of discrete variables or factors in producing particular outcomes. Historians already have to accept the irreducibility of complexity and the humility of interpretation that this entails. Emergence is just a new way to make peace with that intractability.

\section*{NOTES}


10. For one recent documentation of this spread, see Vinod Kataria, “Mobile Phone Use to Rise in Africa, India.” EDN 56.26 (Dec. 15, 2006) 24.


17. This is very similar to the way that Joshua Epstein and Robert Axtell propose to use emergent, agent-based simulations as a corrective to conventional social science. Joshua Epstein and Robert Axtell, Growing Artificial Societies (Boston: MIT Press, 1996).
