Focus: Computational History and Philosophy of Science

Introduction

Abraham Gibson, *Arizona State University*
Manfred D. Laubichler, *Arizona State University and Santa Fe Institute*
Jane Maienschein, *Arizona State University and Marine Biological Laboratory*

Abstract: Digital technologies have transformed both the historical record and the historical profession. This Focus section examines how computational methods have influenced, and will influence, the history of science. The essays discuss the new types of questions and narratives that computational methods enable and the need for better data management in the history and philosophy of science (HPS) community. They showcase various methodological approaches, including textual and network analyses, and they place the computational turn in historiographical and societal context. Rather than surrendering to either technophilia or technophobia, the essays articulate both the benefits and the drawbacks of computational HPS. They agree that the future of the field depends on the successful integration of technological developments, social practices, and infrastructural support and that historians of science must learn to embrace collaboration both within and beyond disciplinary boundaries.

Everyone agrees that historians are living through a revolution, but no one can agree on what we should call it. Over the past several years, *Isis* has published scattered references to "the computational turn," "the computational revolution," and "the electronic information

Abraham Gibson is a National Science Foundation Postdoctoral Fellow in the Center for Biology and Society at Arizona State University. His research explores the history of biology, the ethics of engineering, and the social impact of data science. He recently published *Feral Animals in the American South: An Evolutionary History* (Cambridge, 2016), and he is now writing a book about scientific holism during the interwar period. Center for Biology and Society, Arizona State University, 427 East Tyler Mall, Tempe, Arizona 85287-4501, USA; abraham.gibson@asu.edu.
Manfred D. Laubichler is President's Professor of Theoretical Biology and History of Biology and Director of the Global Biosocial Complexity Initiative at Arizona State University and a Professor at the Santa Fe Institute. His main research question is how something genuinely novel can emerge in evolution and history. School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA; Manfred.Laubichler@asu.edu.
Jane Maienschein is University Professor, Regents Professor, and President's Professor at Arizona State University, where she directs the Center for Biology and Society. She is also a Fellow at the Marine Biological Laboratory in Woods Hole, Massachusetts. Her research explores the history and philosophy of developmental biology; her books include *Embryos under the Microscope: Diverging Meanings of Life* (Harvard, 2014). At the MBL, she codirects a project on understanding regeneration across the scales of life. School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA; maienschein@asu.edu.
This Focus section was organized by Abraham Gibson, Manfred D. Laubichler, and Jane Maienschein.

Isis, volume 110, number 3. © 2019 by The History of Science Society.
All rights reserved. 0021-1753/2019/0110-0004$10.00.

revolution." Two years ago, the annual issue of *Osiris* was dedicated to "the big data revolution," while the *American Historical Review* has recently heralded both "the *digital* revolution" and "the *digitized* revolution." Whatever we call it, there is a growing awareness that digital sources and computational tools have transformed how we engage with the historical record, including the history of science.[1]

Furthermore, these internal discussions are embedded within much larger trends involving rapidly increasing transformations in artificial intelligence, machine learning, and automation that have in a relatively short period affected all aspects of life, including science and education (see also "Computational History of Knowledge: Challenges and Opportunities," in this Focus section). The effects of these transformations are now the subject of critical analysis and will be a major topic for the historical sciences in the future.

The revolution has already produced undeniable benefits for the historical profession. They say that half of knowledge is knowing where to find knowledge—which is to say knowing how to use Google, JSTOR, and, yes, Wikipedia. Entire archives have been digitized, allowing historians to formulate new "grand narratives."[2] They invariably do so on word processors rather than typewriters. These tools and resources influence the history of science just as surely as every other field. The collected papers of Charles Darwin, Albert Einstein, and Isaac Newton have all been digitized and posted online. Like the Internet Archive and HathiTrust, the Biodiversity Heritage Library offers access to millions of pages of keyword-searchable data. Like it or not, Twitter is one of the best ways to connect with fellow historians of science and, increasingly, one's only hope of following now-daily developments in the field.

Not everything is perfect. As in all revolutions, the excitement is tinged with uncertainty. It is true that computational methods and digital resources can help elucidate overlooked peoples and stories, but they can also help reify existing power structures. Uneven digitization risks perpetuating ethnocentrism. Digital media too easily divorce history from its context, a complaint with special significance in our so-called Post-Truth moment. Moreover, it is not at all clear that historians of science are any better prepared than the rest of society to meet the challenges that deepfake technology poses for our evidentiary base. Algorithms dominate every step of the historian's craft—and, increasingly, every aspect of modern life—and they largely do so under the guise of objectivity. And, as for society at large, a major impediment to developing a critical understanding of the challenges and opportunities afforded by these technological transformations is the lack of deep knowledge about the technical aspects of these technologies. This is, of course, primarily a challenge for education and professional training. Without deeper exposure to the theoretical and applied dimensions of this revolution, we will continue to see a bimodal response: either naive acceptance of technological wonders (including the convenience they can bring) or equally naive rejection.

Within the history of science, the computational turn shares basic similarities with other historiographical touchstones. For example, they all reflect broader trends in society and culture. George Sarton's belief that historians of science should take science and its technical aspects seriously encapsulated a broader commitment to objectivity during the early twentieth century.

---

[1] Anna-Luna Post and Andreas Weber, "Notes on the Reviewing of Learned Websites, Digital Resources, and Tools," *Isis*, 2018, *109*:796–800, esp. p. 796; Manfred D. Laubichler, Jane Maienschein, and Jürgen Renn, "Computational Perspectives in the History of Science: To the Memory of Peter Damerow," *ibid.*, 2013, *104*:119–130, esp. p. 119; Stephen P. Weldon, "Introduction [Focus: Ordering the Discipline: Classification in the History of Science]," *ibid.*, pp. 537–539, esp. p. 538; Elena Aronova, Christine von Oertzen, and David Sepkoski, "Introduction: Historicizing Big Data," *Osiris*, 2017, N.S., 32:1–17, esp. p. 2; and Lara Putnam, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast," *American Historical Review*, 2016, *121*:377–402, esp. p. 389.

[2] Post and Weber, "Notes on the Reviewing of Learned Websites, Digital Resources, and Tools," p. 796.

Thomas Kuhn's emphasis on the social aspects of the scientific enterprise echoed an increase in social activism all around the world. The profession's subsequent transition from the history of scientific ideas to the history of scientific practices (à la Steven Shapin, Bruno Latour, and Donna Haraway) coincided with the culture wars of the 1990s.[3] We have clearly arrived at another moment. Technology has transformed every aspect of the human experience, from interpersonal communication to international diplomacy. Our profession's embrace of digital methods cannot be divorced from this context.

And yet, despite its similarity to previous historiographical pivots, the computational turn is in many ways unique. On the one hand, it is pervasive. We may not invoke Sarton or Kuhn every time we write a paper or give a lecture, but we always use a computer, and we rely on the internet for an increasing fraction of the information we process. On the other hand, computational history is far more niche than previous historiographical turns. Few historians know how to read or write code, and almost no history programs teach these skills. The learning curve is formidable indeed. If we want to remain relevant authorities on the past, we will have to learn the basics of computer science or outsource part of our responsibilities to those who have. In the meantime, the emphasis on methods has generally inhibited any larger discussion about theoretical integration, allowing technophilia and technophobia both to run wild. Some insist that computers will remain a supplemental tool in the historian's expanding repertoire. Others insist that computers will eventually render us obsolete.

In this Focus section, we examine the use of computational methods in the history and philosophy of knowledge. We were motivated to organize these papers by recent developments in the field. Over the past few years, computational methods have allowed researchers to examine old questions from new perspectives and to ask new types of questions that previous generations would not have thought possible. In the first essay, on the challenges and opportunities of computational history of knowledge, Manfred Laubichler, Jane Maienschein, and Jürgen Renn describe the types of questions that one can ask using computational methods. This allows historians of science to revisit, and in some cases to settle, old debates. It also allows them to ask entirely new types of questions.

This essay argues that computational history both continues and disrupts trends within the history of knowledge. It continues recent developments that focus on comparisons, *longue durée* accounts, global perspectives, and attempts to grapple with the explosion of size over the last decades. For these historiographical concerns, computational history offers tools that support and, in many cases, enable new narratives. Without computational methods it is simply impossible to process the amount of data needed to capture the dynamics of current science or to map all the relevant events that make a global history of knowledge so challenging and interesting, especially when we consider long historical periods. As an aside, we note that these analyses are, of course, possible only because of a centuries-old tradition of curating archives and libraries. But then, the first computers were also human (and mainly female).

The disruptive aspect of computational history is mainly epistemological in nature. As we process larger and larger datasets and deploy machine learning and related algorithmic methods, we are faced with serious methodological challenges: What training sets do we use? How can we detect relevant changes in patterns and dynamics? How do we interpret correlations and patterns that are not always easy to accommodate within our standard conceptual structures for the evolution of knowledge? How can we infer underlying dynamics from observed patterns? Before we can integrate these analytic results into narratives, we need to assess their validity, and this requires elaborate statistical analysis. History needs to develop frameworks to

[3] Aronova *et al.*, "Introduction" (cit. n. 1), p. 5.

integrate such quantitative results into its inherent narrative structure based on human experience and meaning.

We see this as an important challenge and opportunity, though it will certainly be disruptive. If computational history is going to reach its full potential, historians must learn to manage their data according to more exacting standards. In the second essay, "The Hitchhiker's Guide to Data in the History of Science," Julia Damerow and Dirk Wintergrün discuss the creation and life cycle of a "data corpus" in history of science projects, and they offer a standardized workflow that covers the basics of data management, from planning to acquisition to analysis. They write that historians of science need to start treating their curated data corpora like first-class citizens. Various initiatives at the Max Planck Institute for the History of Science demonstrate the feasibility of such an approach.

One of the main lessons of this essay is the need to bring together technical developments, infrastructure, and social practices. Digital and computational approaches will really work only if they are built on accepted standards and communities of practice. The challenges here are manifold: How can such communities be established and sustained? What are their incentive structures? And how can they be built in such a way that they remain transparent and effective, especially in light of competition from big commercial platforms that often own and control data? While these are issues relevant for all of academia, smaller players, such as the humanities, are especially vulnerable. But they also have an advantage because they have not yet built up too many structures that are difficult to change. This essay suggests ways to begin this process.

The next two essays describe two different kinds of computational methods. In the first, "Triangulation of History Using Textual Data," Ken Aiello and Michael Simeone examine the use of textual analyses in the history of science. They explain that textual analyses can be used to analyze words, phrases, and documents systematically, looking for trends related to meaning and intentions over time. By identifying and analyzing changes in semantic content, historians can ask questions related to the evolution of language, the use and abuse of concepts, and the construction of knowledge. Textual analyses also provide insight into the behavior and specific actions taken by individuals, social groups, national economies, and global trade networks.

Aiello and Simeone note that large volumes of data allow historians to ask questions and analyze sources across multiple scales. The challenge is then how to map the results from analyses at different scales onto each other in order to arrive at a truly integrated multiscale interpretation. This also means that we need to be able to combine different kinds of data, such as textual and social data. This essay shows how such data integration can work and demonstrates what novel insights into the social and linguistic context of historical events can be derived from such analyses.

In "Historical Network Analysis of Evolutionary Medicine," Deryc Painter, Bryan C. Daniels, and Jürgen Jost illustrate the diversity of insights that can be derived from a well-constructed corpus using network-based approaches. Complex networks are a main method for representing large connected datasets. They have been used in the social sciences, and to a lesser degree in the humanities, for some time. The main challenge pertaining to network analysis has been to find the right formal measures for meaningful historical insights. All too often network visualizations are used as a sort of token, intended to sway the reader without providing concrete insights. This essay discusses some basic and some more advanced formal aspects of networks that can be connected to concrete historical questions and illustrates these with concrete examples from a case study on evolutionary medicine.

A challenge for network approaches is the need to have well-curated datasets. Here the authors give some guidelines for how to build a corpus and prepare it for network analysis. This essay also illustrates the sorts of novel collaborations that characterize computational history of knowledge. The authors are not seeking just to use a few standard metrics that are easy to calculate; rather, they hope to develop more adequate formal approaches that actually measure features of the network and have the potential to contribute insights with regard to historical questions. These are benefits to be realized from the collaboration of mathematicians (in this case, Jost), physicists (Daniels), and historians (Painter).

In the final essay, "The History of Science and the Science of History: Computational Methods, Algorithms, and the Future of the Field," Abraham Gibson and Cindy Ermus attempt to place the computational turn in historiographical and societal context. They ask whether comparisons to quantitative history are warranted, whether historians with no previous training can learn and apply computational methods to any project in the history of science, and what it all means for the profession in general. In particular, following a generalized workflow that was developed in the Laubichler Lab for Computational HPS, the authors used a variety of tools to help them analyze and visualize biological theories about cooperation during the interwar period.

Gibson and Ermus confirm that the digital revolution has had many positive effects, including the democratization of knowledge and the diversification of the historical profession. Digital projects invariably require collaboration, sometimes within the discipline but more often across disciplinary boundaries. This allows historians to consult diverse kinds of evidence and to access new perspectives within the historical record. Even so, the same revolution poses unprecedented challenges. Some complain that digital methods divorce history from its context, that uneven digitization initiatives help reify existing narratives, and that we are beholden to algorithms that we had no hand in crafting and that we largely do not understand. The authors insist that we must demand a place for humans in the digital humanities.

We will be the first to acknowledge that our contributions look different than traditional histories of science. Most obviously, there are a lot of us. Over the past ten years, this journal's Focus sections have averaged a total of 5.9 authors. By comparison, there are 11 of us. During that same span, articles averaged 1.1 authors each. We average 2.5 authors per article, a fact that reflects the collaborative nature of the digital humanities. We are also diverse. This collection features contributions from historians, philosophers, sociologists, biologists, a physicist, and computer scientists. We represent a variety of races, nationalities, and genders.

Obviously, we are excited by the possibilities of computational history and philosophy of knowledge. And by placing these methodological and epistemological developments within a lineage of similar historiographical transformations, we certainly flirt with the label of being "revolutionary." But we are also well aware of the story of Icarus and of some hard facts about the evolution of technology, which is full of promising inventions that never became innovations because they did not connect with what people and markets found useful. The road ahead will thus be challenging and, we hope, rewarding as well.