Calculating Devices and Computers

Matthew L. Jones

PREPRINT
Final version to appear in
Blackwell Companion to the History of Science ed. Bernie Lightman

Abstract:

Focusing upon computation, storage, and infrastructures for data from the early modern European period forward, this chapter stresses that the constraints of computing technologies, as well as their possibilities, are essential for the path of computational sciences. Mathematical tables and simple contrivances aided calculation well into the middle of the twentieth century. Digital machines replaced them slowly: adopting electronic digital computers for scientific work demanded creative responses to the limits of technologies of computation, storage, and communication. Transforming the evidence of existing scientific domains into data computable and storable in electronic form challenged ontology and practice alike. The ideational history of computing should pay close attention to its materiality and social forms, and the materialist history of computing must pay attention to its algorithmic ingenuity in the face of material constraints.

Keywords: calculator, computer, information technology, data, database, approximation, numerical analysis, simulation, expert knowledge

Trumpeting the dramatic effects of terabytes of data on science, a breathless *Wired* article from 2008 described "a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear." No more theory-laden era: "Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves" (Anderson 2008; see Leonelli 2014; Strasser 2012). A new empirical epoch has arrived.

Such big data positivism is neither the first, nor the last, time that developments in information technology have been seen as primed to upset all the sciences simultaneously. Rarely have digital computers and claims of their revolutionary import been far apart. This chapter explores the machines, mathematical development, and infrastructures that make such claims thinkable, however historically and philosophically unsustainable. The chapter focuses upon computation, storage, and infrastructures from the early modern European period forward.[1] A remarkable self-reflexive approach to the very limits of computational tools has long been central to the productive quality of these technologies. Whatever the extent of computational hubris, much generative work within the computational sciences rests on creative responses to the limits of technologies of computation, storage, and communication. Scientific computation works within a clear eschatology: the promised land of adequate speed and storage are ever on the horizon, but, in the meanwhile, we pilgrims in this material state must contend with the materialities of the here and now.

However revolutionary in appearance, the introduction of electronic digital computers as processors, as storage tools, and a means of communication often rested initially upon existing practices of computation and routinization of data collection, processing, and analysis, in science and industry alike.[2] But just as computing altered the sciences, the demands of various sciences altered scientific computing. Transforming the evidence of existing scientific domains into data computable and storable in electronic form challenged ontology and practice alike; it likewise demanded different forms of hardware and software. If computers could tackle problems of such complexity that would otherwise prove infeasible, if not intractable, they did so through powerful techniques of simplification, through approximation, through probabilistic modeling, through means for discarding data and many features of the data.

This chapter focuses upon computational science and science using information technology, rather than the discipline of computer science. Centered on computation (arithmetical operations, integration) and data storage and retrieval, first in Europe, then primarily in the U.S., it omits the story of networks and the Internet, and the role of computational metaphors and ontologies within the sciences.[3] The approach here is episodic and historiographic, rather than comprehensive or narrative.

The constraints of computing technologies, and not just their possibilities, are essential for the path of computational sciences in recent years. To borrow a modish term, we need to give more epistemic attention to the *affordances* of different systems of calculation. The ideational history of computing must thus pay close attention to its materiality and social forms, and the materialist history of computing must pay attention to its algorithmic ingenuity in the face of material constraints.

Calculation "by hand"

The first detailed publication concerning an electronic digital computer appeared in the prestigious journal *Mathematical Tables and other Aids to Computation*, sponsored by no less than the US National Academy of Sciences (Polachek 1995). The first issues of this revealingly named publication in 1943 included a detailed review of basic mathematical tables of logarithms, trigonometric functions, and so forth. The spread of mechanical calculating machines from the late nineteenth century had made the production of tables more, not less, important. "As calculating machines came into use, the need for seven-place tables of the natural values of the trigonometric functions stimulated a number of authors to prepare them" (C[omrie] 1943, 8). Beneath the surface, however, the calculations behind these tables were of surprising antiquity. The foremost major advocate for scientific computation using mechanical calculating machines in the early twentieth century, Leslie Comrie, argued that careful examination of the errors in tables strongly indicated that most of these new sets of them simply were taken, usually with no attribution, from sixteenth- and seventeenth-century tables.[4]

The upsurge of mathematical astronomy in early modern Europe, most associated with Nicolas Copernicus, Tycho Brahe, and Johannes Kepler, spurred the development of new methods for performing laborious calculations, particularly techniques for abridging multiplication by some sort of reduction to addition (Thoren 1988). While the first widespread techniques involved trigonometric identities, John Napier devised the more straightforward technique of logarithms early in the seventeenth century. With logarithms, multiplication and division reduced to addition and subtraction. Put into a more elegant as well as base ten form by Napier's collaborator, the English mathematician Henry Briggs, logarithms soon became a dominant tool in astronomical and scientific calculation well into the twentieth century (Jagger 2003). "Briggs' industry," Comrie explained, "in tabulating logarithms of numbers and of trigonometrical functions made Napier's discovery immediately available to all computers"—that is, people performing calculations (C[omrie] 1946, 149). So basic was logarithmic computation that tables of functions by and large provided logarithms of those functions, rather than regular values, well into the first half of the twentieth century.

Far from replacing tables, mechanical calculating machines gained their currency within scientific applications largely by abridging the labor of the additions of results taken from the tables. One tool complimented the other. The first known mechanical digital calculating machine, that of Kepler's correspondent Wilhelm Schickard, was designed precisely to ameliorate the addition of results taken from Napier's bones. In the mid seventeenth century Blaise Pascal and Gottfried Leibniz envisioned machines to aid financial and astronomical calculation. Despite decades of work Leibniz never brought to any sort of completion his envisioned machine for performing addition, subtraction, multiplication and division directly. And despite a long process of invention and re-invention throughout the eighteenth century, mechanical calculating machines had not become robust enough for everyday financial or scientific use.[5] "In the present state of numerical science," a learned reviewer remarked in 1832, "the operations of arithmetic may all be performed with greater certainty and dispatch by the common method of computing by figures, than almost by any mechanical contrivance whatsoever." More manual devices were far more significant:

> we must except the scale and compasses, the sector, and the various modifications of the logarithmic line with sliders, all of which are valuable instruments. . . . The chief excellence of these instruments consists in their simplicity, the smallness of their size, and the extreme facility with which they may be used.

In sharp contrast, these "qualities which do not belong to the more complicated arithmetical machines, and which . . . render the latter totally unfit for common purposes" (Adam 1832, 400). Mechanical calculating machines entered into wide use, notably among actuaries and accountants, in Western Europe and the United States only in the 1870s at the earliest—and not without continuing skepticism.[6] Simpler mechanical devices, above all the slide-rule, a device using logarithms, remained important into the 1970s.

Charles Babbage envisioned his Difference Engine early in the nineteenth century just to produce mathematical tables in an automated way: the Engine was to be a machine for automating the production of the paper tools central to scientific and business calculation. Babbage sought to ameliorate two aspects of table-making: the calculation of the values and, nearly as important, their typesetting (Swade 2001; Schaffer 1994). Although none of Babbage machines were completed, others, notably the Swedish son and father Scheutz team, produced a working device that saw some use (Lindgren 1990). Securing the order of calculation meant securing the printing process. Problems with print greatly worried all those concerned with scientific computation well into the mid twentieth century.

Mechanical contrivances were not limited to arithmetical operations. Initially developed for aiding census taking, tabulating equipment soon became central to the data intensive life insurance market. Life insurance firms pushed the corporations selling tabulators to develop and refine these machines, to encompass printing, automatic control, sorting and the introduction of non-numerical, alphabetical data (Yates 1993). They offered a materialization of data processing at a large scale that had been brought to a very high level of reliability by the 1920s. At that time, they began to be used for scientific computation in larger numbers (Priestley 2011, ch. 3). Two major advocates, Comrie, and his American analogue, Wallace Eckert, preached the virtues of connecting two largely independent traditions of business machines: calculating machines and register machines, capable of arithmetical operations, and tabulating machines, capable of recording and reading large amounts of data.

Calculation "by hand" did not exclusively comprise manual arithmetic. Calculation by hand encompassed an array of techniques and tools aiding computation, from slide rules to mechanical calculators, and especially mathematical tables of important mathematical functions (Kidwell 1990). And it often involved teams of human calculators, in many cases groups of women (Grier 2005; Light 1999). Well after the advent of electronic digital machines following World War II, scientists in the U.S. and U.K. weighed the costs and benefits of using teams of human computers and punch-card tabulators rather than expensive and hard to access electronic computers (Chadarevian 2002, 111–118).

Analog computing

In 1946, Leslie Comrie remarked,

> I have sometimes felt that physicists and engineers are too prone to ask themselves 'What physical, mechanical or electrical analogue can I find to the equation I have to solve?' and rush to the drawing board and lathe before enquiring whether any of the many machines that can be purchased over the counter will not do the job.

Comrie was decrying a rich tradition of building highly specialized devices that served as physical analogues allowing the solution to problems not otherwise tractable (Care 2006; Mindell 2002; Owens 1986). Such computers were "analog" in two senses: they measured continuous quantities directly and they were "analogous" to other physical phenomena. The best known of these machines, exemplified by Vannevar Bush's differential analyzer, allowed for mechanical integration, and thus were important in the solution of differential equations. Rather than an exact, analytical solution using highly simplified equations, mechanical integrators promised approximate solutions to problems in their much fuller complexity. The superiority of analog computation to numerical approximation for many purposes was still felt in 1946. Praising Bush's differential analyzer, Comrie noted, "Although differential equations can be (and are) solved by finite difference methods on existing machines, the quantity of low-accuracy solutions required today is such that time and cost would be prohibitive. The use of machines for handling infinitesimals rather than finite quantities has fully justified itself…" (C[omrie] 1946, 150). Although digital electronic computers soon eclipsed analogue computers, they did so in many cases less by explicitly solving numerical problems, as by simulating them—a new form of analogical reasoning.

Electronic computing, numerical analysis, and simulation

The demands of war, first World War II and then the early Cold War, provided impetus and funding alike for the development of electronic digital computing in the United States, Britain, and the Soviet Union.[7] In 1946, John von Neumann and his collaborator H. H. Goldstine declared, "many branches of both pure and applied mathematics are in a great need of computing instruments to break the present stalemate created by the failure of the purely analytical approach to non-linear problems" (Von Neumann and Goldstine 1961, 4; Dahan Dalmenico 1996, 175). Working with electronic computers meant recognizing their affordances and limits. The "computing sheets of a long and complicated calculation in a human computing establishment" can store more than all the new electronic computers. They concluded,

> … in an automatic computing establishment there will be a 'lower price' on arithmetical operations, but a 'higher price' on storage of data, intermediate results, etc. Consequentially, the 'inner economy' of such an establishment will be very different from we are used to now, and what we were uniformly used to since the day Gauss. . . . new criteria for 'practicality' and 'elegance' will have to be developed. . . (Von Neumann and Goldstine 1961, 6; Aspray 1989, 307–8)

The new electronic digital computers produced just after World War II offered great possibility for speedy computation, while demanding their users rework older methods of numerical analysis. In the context of work around the atomic bomb, von Neumann altered numerical methods for solving partial differential equations in fluid dynamics better to allow them to be digitally calculated. Modifying existing approaches to numerical analysis to comport with this "inner economy," von Neumann and others spurred the development of new numerical analyses ever more tailored for the constraints and power of digital electronic machines, in particular the challenges of round-off error.

As the science of computerized numerical analysis developed, its limits became ever more clear, especially in the context of designing thermonuclear weapons (Galison 1997 ch. 8). Before the war, the physicist Enrico Fermi had worked on the idea of creating mathematical simulations of atomic phenomena. Stansilaw Ulam, along with von Neumann and Nicolas Metropolis, devised an approach dubbed "Monte Carlo" to tackle the challenging problems of studying the interactions with a nuclear weapon. The idea was to sample a large set of simulated outcomes of a process or situation, rather than attempting to solve analytically, or even numerically, the differential equations governing the process. Ulam began with the game of solitaire. One could generate a large number of different solitaire games, without enumerating them all, then analyze statistically the properties of that set of games. The same sort of analysis could be applied to the study of nuclear phenomena. Such simulations, remarkably, worked for many classes of problems without any stochastic content, such as the solution of an integral or the value of π. Something currently intractable theoretically became quasi-experimental. As Ulam and Metropolis noted, the potency of Monte Carlo came just because it could sidestep computationally intractable problems:

> The essential feature of the process is that we avoid dealing with multiple integrations or multiplications of the probability matrices, but instead sample single chains of events. We obtain a sample of the set of all such possible chains, and on it we can make a statistical study of both the genealogical properties and various distributions at a given time (Metropolis and Ulam 1949, 339).

Monte Carlo and other such simulations rested then on a critique of human and artificial reasoning.[8]

Monte Carlo heralded the emergence of simulation as a central form of scientific knowledge in the years following the war (Seidel 1998; Galison 1997, 779; Lenhard, Shinn, and Küppers 2006) Computer simulation provided a novel sort of science sitting uncomfortably between experiment and theory and required a dramatic reconfiguration of adequate scientific knowledge. This reconfiguration was in many cases bitterly resisted, before becoming naturalized and now a central aspect of scientific practice. Originally used to sidestep the intractability of differential equations, simulations now come in many forms. Some generate simulations using underlying theoretical models; others eschew any claim to represented underlying theoretical structure and aim simply at behavioral reproduction. As so often in the history of science, the lack of closure about the philosophical issues around such a transformation has not precluded widespread adoption of the approach. Indeed, that lack of closure created the space for the creation of new—if often tendentious—approaches to the study of complex systems without the need for reduction to covering laws and highly simplified models.

Beyond Artillery, Bombs, and Particles

"How could a computer that only handles numbers be of fundamental importance to a subject that is qualitative in nature and deals in descriptive rather than analytic terms?" (as quoted in November 2012, 20). Such a concern, here about biology, was true of numerous domains of knowledge. The success of early electronic computers following the Second World War within traditionally heavily quantitative domains such as atomic physics and ballistics did not make it evident that computers had much to offer to rather different sciences. In field after field, pioneers nevertheless sought to transform the evidence and forms of reasoning of scientific subfields into new, more computationally tractable forms. The adoption of computing was neither natural or easy (Yood 2013). In his recent history of biological computing, Hallam Stevens argues against the contention that superior computer power and storage capacity allowed biologists finally to adopt computerized tools in great number. Instead, he argues, biology "changed to become a computerized and computerizable discipline" (Stevens 2013, 13). Even in highly quantitative domains, the means for rendering problems appropriate to computation came from an array of disciplines, many initially created in wartime work: developments in fluid mechanics, statistics, signals processing, and operations research each provided distinctive ways of making problems computationally tractable (Dahan Dalmenico 1996). The plurality of approaches remains marked in the multiple names attached to many roughly similar computational techniques.[9]

For all the recent philosophical and historical work on models and simulations, we have no solid taxonomy of the varied forms of reflective simplification, reduction, and transformations of problem domains so that they become computationally tractable.[10] A great deal of the ingenuity of the application of computers to the sciences comes just in the creative transformation of problem domains conjoined to arguments about the scientific legitimacy of that transformation. As might be expected, reductions and simplifications that were initially bitterly contested became standard practice in subfields, and their contingency was lost. These reductions involved simplifications of data and of underlying possible models alike; they can also involve transformations in what suffices as scientific knowledge. We have a highly ramified set of different mixes of instrumentalism and realism still in need of good taxonomies.

In his commanding study of climate science, for example, Paul Edwards describes the emergence of a new ideal of "reproductionism" within computation of science that "seeks to stimulate a phenomenon, regardless of scale, using whatever combination of theory, data, and 'semi-empirical' parameters may be required." In this form of science, he argues, the "familiar logics of discovery and justification apply only piecemeal. No single, stable logic can justify the many approximations involved in reproductionist science" (Edwards 2010, 281). The line between the empirical and the theoretical has become productively blurred.

Herbert Simon, to take a second example, famously offered a contrast between approaches in operations research and then current artificial intelligence perspectives on decision problems. The algorithms of operations research, he noted, "impose a strong mathematical structure on the decision problem. Their power is bought at the cost of shaping and squeezing the real-world problem to fit their computation: for example, replacing the real-world criterion function and constraint with linear approximation so that linear programming can be used." In contrast, he explained, "AI methods generally find only satisfactory solutions, not optima . . .we must trade off satisficing in a nearly-realistic model (AI) against optimizing in a greatly simplified model (OR)" (Simon 1996, 27–28; November 2012, 274).

These debates have continued into the era of data mining and big data. In 2001, the renegade statistician Leo Breiman polemically described the divide among two major statistical cultures:

> Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** (independent variables) go in one side, and the other side the response variables **y** come out. Inside the black box, nature

> functions to associate the predictor variables with the response variables. . . .
> These are two [distinct] goals in analyzing the data:
> *Prediction*. To be able to predict what the responses are going to be to future input variables;
> *Information*. To extract some information about how nature is associating the response variables to the input variables (Breiman 2001, 199).

Against the dominant statistical view, Breiman argued for an "algorithmic modeling culture" that is satisfied with the goal of prediction without making physical claims about the actual natural processes. Variants of such epistemic modesty are central to much recent work in machine learning, yet many scientists and statisticians find it far too instrumentalist.

Big Data avant Big Data

In the late 1940s, Soviet cryptography abruptly became very strong and largely impervious to decryption by the US and its allies. The signals intelligence agencies of the West, notably the newly established National Security Agency, found themselves early in the Cold War needing the capacity to process large amounts of data far more than the capacity to perform arithmetic quickly. Under the sponsorship of the US national laboratories concerned with nuclear weapons, computer developments focused to a great extent upon improving the processing speed needed for simulations using floating-point arithmetic (MacKenzie 1991, 197). In contrast, the NSA needed to be able to sort through large amounts of traffic quickly: "the Agency became as much or more a data processing center than a 'cryptanalytic center." As a result NSA sought "high speed substitutes for the best data processors of the era, tabulating equipment" (Burke 2002, 264). In focusing "on the manipulation of large volumes of data and great flexibility and variety in non-numerical logical processes,” NSA had needs more akin to most businesses than to physicists running simulations. Just as substantial federal funds promoted the creation of ever faster arithmetical machines, substantial federal funds for cryptography sponsored intense work on larger storage mechanisms. The two came together, with great friction, in funding IBM’s attempts to create a jump in capability in the mid 1950s. “AEC’s computer requirement emphasized high-speed multiplication, whereas NSA’s emphasis was on manipulation of large volumes of data and great flexibility and variety in non-numerical logical processes” (Snyder 1980, 66).

The sciences followed suit. In 1950, Mina Rees of the Naval Research Office noted the “great emphasis” in early machines “that would accept a small amount of information, perform very rapidly extensive operations on this information, and turn out a small amount of information as its answer…” Now, she wrote, the interest “seems to lie in a further exploration of the use of machines to accept large amounts of data, perform very simple operation upon them, and print out, possibly, very large numbers of results” (Rees 1950, 735). The experimental data produced in high-energy physics quickly challenged storage and processing abilities alike (Seidel 1998, 54). In science as in snooping, the data potentially to be analyzed and stored has ever outstripped processing power, memory, and storage capacity. “Over the past 40 years or more,” a piece in *Science* noted in 2009, “Moore’s Law has enabled transistors on silicon chips to get smaller and processors to get faster. At the same time, technology improvements for disks for storage cannot keep up with the ever increasing flood of scientific data generated by the faster computers” (Bell, Hey, and Szalay 2009).

These material constraints challenged scientists and applied mathematicians to develop ways of abridging and reducing data and to account for the legitimacy of such reductions. In an early effort at computing Fourier syntheses for use in X-ray crystallography, J. M. Bennett and J.C. Kendrew explained, "In a machine such as the EDSAC [Electronic Delay Storage Automatic Calculator], . . . , it is impossible to accommodate all the terms of a typical two-, and more, especially of a three-, dimensional synthesis…" The authors devised and defended numerous techniques for representing the data more compactly without losing too much significant information, including "smoothing": ". . . in many cases the synthesis obtained from such smoothed-off data is not significantly different from that compounded of unsmoothed data" (Bennett and Kendrew 1952, 112). Such reflective reductions of the data—with potentially dangerous loss of information—remain integral to nearly all data intensive scientific work, and have only become more central with the petabytes of "big data." Algorithms drawn from statistics, initially developed for smaller data sets, often require dramatic transformation; one 1996 paper, for example, explains that approaches from artificial intelligence and statistics "do not adequately consider the case that the dataset can be too large to fit in main memory. In particular, they do not recognize that the problem must be viewed in terms of how to work with a [*sic*] limited resources (e.g. memory that is typically, much smaller than the size of the dataset) to do the clustering as accurately as possible while keeping I/O [input/output] costs low" (Zhang, Ramakrishnan, and Livny 1996, 104). Large amounts of data thus often become the grounds for a technological deterministic account of the necessity of computational choices. Not just the number of observations, but the dimensionality of those observations requires transformative techniques for reducing and choosing among the data.

Data Infrastructures

Computerized storage of data is not neutral, obvious, or natural.[11] Obtaining data from the world is hard work; standardizing it often more so. Standardizing data is intensive and non-trivial. Contemporary data scientists often quip that some 95 percent of analytical time involves "data munging." This commonplace is borne out in within science studies. In his study of twentieth-century climatology, Paul Edwards discusses the distinct process of "making global data"—collecting weather data from around the world and of "making data global"—"building complete, coherent, and consistent global data sets from incomplete, inconsistent, and heterogeneous data sources" (Edwards 2010, 251).

Standardizing data is challenging for individual scientific research groups, but even more contested while crossing institutional, disciplinary, and national lines. In their study of model organism databases, Sabine Leonelli and Rachel Ankeny describe the development of formalized data curators responsible for "(1) the choice of terminology to classify data and (2) the selection and provision of information about experimental settings in which data are produced, including information about specimens and protocols" (Leonelli and Ankeny 2012, 31). To be useful, datasets must include metadata: information about the production of the data, essential for evaluating it in further analysis. Dispute over the approach can produce what one group calls "science friction" (Edwards et al. 2011).

Each database depends a set of decisions—and compromises—about how to represent which aspects of data collection and how to store them on actual computer systems.[12] These choices at once limit and make possible particular forms of knowledge production. The cataloging of genomes saw the "evolution of GenBank from flat-file to relational to federated data-based," which Hallam Stevens has argued, "paralleled biologists' moves from gene-centric to alignment-centric to multielement views of biological action" (Stevens 2013, 168). Leonelli and Ankeny likewise note, "Through classification systems such as the Gene Ontology, databases foster implicit terminological consensus within model organism communities, thus strengthening communication across disciplines but also imposing epistemic agreement on how to understand and represent biological entities and processes" (Leonelli and Ankeny 2012, 32). The point is not that databases allow only one sort of theory: different databases lend themselves to particular types of investigation and make others more challenging. Different ways of storing data have different investigative affordances. Like models, database can be performative (Bowker 2000, 675–6).

Advocates of the introduction of computation into various scientific fields draw heavily upon technological determinist narratives to justify the necessity of new epistemic practices and differently skilled practitioners. To justify the intrusion of computational statistical methods into taxonomy, for example, the biologist George Gaylor Simpson explained that they "become quite necessary as we gather observations on increasing large numbers of variables in large numbers of individuals" (Simpson 1962, 504).

The Social Organization of Expertise

In 1962, Simpson envisioned new forms of computational taxonomy in zoology:

> the day is upon us when for many of our problems, taxonomic and otherwise, freehand observation and rattling off elementary statistics on desk calculators will no longer suffice. The zoologist of the future, including the taxonomist, often is going to have to work with a mathematical statistician, a programmer, and a large computer. Some of you may welcome this prospect, but others may find it dreadful (Simpson 1962, 504–5; see Hagen 2001).

Practices of computation rest on social organizations of expertise. Debates about the propriety of using calculating tools often hinge on the distribution of skill and boundaries of expertise. Having just advocated the necessity of statistical computing, Simpson defended the continuing necessity of the trained human biologist against "extremists" who "hold that comparison of numerical data on samples by means of a computer automatically indicates the most natural classification of the corresponding populations." While "computer manipulation has become not only extremely useful and indispensible," he explained, it is false that "it can automatically produce a biologically significant taxonomic result" (Simpson 1962, 505).

Such demarcation battles figure prominently in the many sciences computerized in the second half of the twentieth century. Peter Galison documented the conflict within post war microphysics concerning the necessity of human interpretation of high-energy events. Committed to the discovery of novel, startling events, the physicist Luis Alvarez stressed the distinctiveness of human cognitive capacities. Insisting on a "strong positive feeling that human beings have remarkable inherent scanning abilities," Alvarez declared, "these feelings should be used because they are better than anything that can be built into a computer" (as quoted in Galison 1997, 406). Attendant upon this epistemic claim was the need for an industrial organization of human scanners possessing such feelings.

Programming—or teaching—computer to perform acts of judgment and inference motivated major work in artificial intelligence. Notable successes included attempts to formalize the judgment of scientists concerning organic chemical structures, as in the case of the expert system DENDRAL (November 2012, 259–268). By the early 1970s, many practitioners worried greatly about the challenge of converting human expertise into "knowledge-bases" and formal inference rule. In a move akin to Harry Collins' reinvigoration of "tacit knowledge" in the sociology of science, artificial intelligence researchers became worried about the "knowledge acquisition bottleneck" (Edward Feigenbaum 2007, 62–63; Forsythe 1993). J. Ross Quinlan noted that part "of the bottleneck is perhaps due to the fact that the expert is called upon to perform tasks that he does not ordinarily do, such as setting down a comprehensive roadmap of a subject" (Quinlan 1979, 168). Rather than attempting to simulate some aspect of the cognitive process of judgment, new forms of pattern recognition and machine learning attempted to predict the expert judgments based on the behavior of experts in some task of classification. ". . . the machine learning technique takes advantage of the data and avoids the knowledge acquisition bottleneck by extracting classification rules directly from data. Rather than asking an expert for domain knowledge, a machine learning algorithm observes expert tasks and induces rule emulating expert decisions" (Irani et al. 1993, 41). Just such a positivist dream about the possibilities of such instrumentalist learning algorithms ultimately inspired the breathless *Wired* article with which I began.

While attempts to automate aspects of human cognition inspired machine learning, another strand of research sought to optimize computer output best to draw upon human potential. A National Science Foundation sponsored report in 1987 noted, the “gigabit bandwidth of the eye/visual cortex system permits much faster perception of geometric and spatial relationship than any other mode, making the power of supercomputers more accessible.” The goal was to harness the brain, not sidestep it. “The most exiting potential of wide-spread availability of visualization tool is … the insight gained the mistakes understood by spotting visual anomalies while computing. Visualization will put the scientist into the computing loop and change the way science is done” (McCormick, DeFanti, and Brown 1987, vii, 6). A celebration of embodied minds, scientific visualization brought together the affordances and limits of human beings and machines alike.[13]

Hubris and Materiality

A 2006 piece in *Science* located the coming of a new data-focused science within a classical narrative of the history of science:

> Since at least Newton’s laws of motion in the 17th century, scientists have recognized experimental and theoretical science as the basic research paradigms for understanding nature. In recent decades, computer simulations have become an essential third paradigm: a standard tool for scientists to explore domains that are inaccessible to theory and experiment, such as the evolution of the universe, car passenger crash testing, and predicting climate change.

Information systems, the authors claims, have now moved beyond simulation:

> As simulations and experiments yield ever more data, a fourth paradigm is emerging, consisting of the techniques and technologies needed to perform data-intensive science . . .

And yet this prophecy of a coming age lacks eschatological vim; its concerns are infrastructural and material. The vast data now available outstrips storage, processing, and communications resources. “In almost every laboratory, ‘born digital’ data proliferate in files, spreadsheets, or databases stored on hard drives, digital notebooks, Web sites, blogs, and wikis. The management, curation, and archiving of these digital data are becoming increasingly burdensome for research scientists." The problem rests on a lack of understanding of the material conditions for data-intensive science: "data-intensive science has been slow to develop due to the subtleties of databases, schemas, and ontologies, and a general lack of understanding of these topics by the scientific community." Too ideational a conception of computational science, in other words, has slowed the development of a data-driven computation science: "In the future, the rapidity with which any given discipline advances is likely to depend on how well the community acquires the necessary expertise in database, workflow management, visualization, and cloud computing technologies." (Bell, Hey, and Szalay 2009, 1297-8)

Devices for computing and information storage have long challenged their users: far from leading users into a virtual world without the challenges of the material one, they require their users to contend with their affordances and material limits. These limits—in processing power, in storage size and speed, in bandwidth—demand much of users, and users have done much with them.

References

Adam, Anderson. 1832. "Arithmetic." In *The Edinburgh Encyclopaedia Conducted by David Brewster, with the Assistance of Gentlemen Eminent in Science and Literature*, edited by David Brewster, 2:345–400. J. and E. Parker.

Agar, Jon. 2003. *The Government Machine: A Revolutionary History of the Computer*. Cambridge, Mass.: MIT Press.

———. 2006. "What Difference Did Computers Make?" *Social Studies of Science* 36, No. 6: 869–907. doi:10.1177/0306312706073450.

Akera, Atsushi. 2007. *Calculating a Natural World: Scientists, Engineers, and Computers During the Rise of U.S. Cold War Research*. Cambridge, Mass.: MIT Press.

Anderson, Chris. 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” *Wired Magazine On-Line*. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.
Aspray, William. 1989. “The Transformation of Numerical Analysis by the Computer: An Example from the Work of John von Neumann.” In *History of Modern Mathematics*, edited by David E. Rowe and John McCleary, Academic Press, 2:307–22. Boston.
———. , ed. 1990. *Computing before Computers*. Ames: Iowa State University Press.
Bell, Gordon, Tony Hey, and Alex Szalay. 2009. “Beyond the Data Deluge.” *Science* 323, No. 5919: 1297–98.
Bennett, John M., and John C. Kendrew. 1952. “The Computation of Fourier Synthesis with a Digital Electronic Calculating Machine.” *Acta Crystallographica* 5, No. 1: 109–16.
Bergin, Thomas J., and Thomas Haigh. 2009. “The Commercialization of Database Management Systems, 1969–1983.” *Annals of the History of Computing* 31, No. 4: 26–41.
Bowker, Geoffrey C. 2000. “Biodiversity Datadiversity.” *Social Studies of Science* 30, No. 5: 643–83.
Breiman, Leo. 2001. “Statistical Modeling: The Two Cultures.” *Statistical Science* 16: 199–215.
Brezinski, C., and L. Wuytack. 2001. “Numerical Analysis in the Twentieth Century.” In *Numerical Analysis: Historical Developments in the 20th Century*, edited by L. Wuytack and C. Brezinski, 1–40. Amsterdam: Elsevier. http://www.sciencedirect.com/science/article/pii/B9780444506177500033.
Burke, Colin B. 2002. *It Wasn’t All Magic: The Early Struggle to Automate Cryptanalysis, 1930s-1960s*. Fort Meade, MD: Center for Cryptological History, NSA. http://archive.org/details/NSA-WasntAllMagic_2002.
Burri, Regula, and Joe Dumit. 2008. “Social Studies of Scientific Imaging and Visualization.” In *The Handbook of Science and Technology Studies*, edited by Edward J. Hackett, 3rd ed., 297–317. Cambridge, MA: The MIT Press.
Care, Charles. 2006. “A Chronology of Analogue Computing.” *The Rutherford Journal: The New Zealand Journal for the History and Philosophy of Science and Technology* 2, No. July. http://www.rutherfordjournal.org/article020106.html.
Chadarevian, Soraya de. 2002. *Designs for Life: Molecular Biology After World War Ii*. Cambridge: Cambridge University Press.
C[omrie], L. J. 1943. “Recent Mathematical Tables.” *Mathematical Tables and Other Aids to Computation* 1, No. 1: 3–23. doi:10.2307/2002683.
———. 1946. “The Application of Commercial Calculating Machines to Scientific Computing.” *Mathematical Tables and Other Aids to Computation* 2, No. 16: 149–59. doi:10.2307/2002577.
Cortada, James W. 2000. *Before the Computer: IBM, NCR, Burroughs, and Remington Rand and the Industry They Created, 1865-1956*. Princeton, N.J.: Princeton University Press.
———. 2012. *The Digital Flood: The Diffusion of Information Technology across the U.S., Europe, and Asia*. New York: Oxford University Press.

Creager, Angela N. H., Elizabeth Lunbeck, and M. Norton Wise, eds. 2007. *Science Without Laws: Model Systems, Cases, Exemplary Narratives*. Durham: Duke University Press.

Crowe, G.D., and S.E. Goodman. 1994. "S.A. Lebedev and the Birth of Soviet Computing." *Annals of the History of Computing, IEEE* 16, No. 1: 4–24. doi:10.1109/85.251852.

Dahan Dalmenico, Amy. 1996. "L'essor des mathématiques appliquées aux États-Unis: l'impact de la seconde guerre mondiale." *Revue d'histoire des mathématiques* 2, No. 2: 149–213.

Edward Feigenbaum. 2007. Oral History of Edward Feigenbaum Interview by Nils Nilsson. http://archive.computerhistory.org/resources/access/text/2013/05/102702002-05-01-acc.pdf.

Edwards, Paul. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge: MIT Press.

Edwards, Paul, Matthew S. Mayernik, Archer Batcheller, Geoffrey Bowker, and Christine Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." *Social Studies of Science* 41, No. 5: 667–90.

Forsythe, D. E. 1993. "Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence." *Social Studies of Science* 23, No. 3: 445–77. doi:10.1177/030631279302300300 2.

Galison, P. 1997. *Image and Logic: A Material Culture of Microphysics*. University of Chicago Press.

Goldstine, Herman H. 1972. *The Computer from Pascal to von Neumann*. Princeton, N.J. : Princeton University Press.

Goodman, Seymour. 2003. "The Origins of Digital Computing in Europe." *Communications of the ACM* 46, No. 9: 21–25.

Grier, David Alan. 2005. *When Computers Were Human*. Princeton: Princeton University Press.

Hagen, Joel B. 2001. "The Introduction of Computers into Systematic Research in the United States during the 1960s." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 32, No. 2: 291–314. doi:10.1016/S1369-8486(01)00005-X.

Haigh, Thomas. 2009. "How Data Got Its Base: Information Storage Software in the 1950s and 1960s." *Annals of the History of Computing, IEEE* 31, No. 4: 6–25.

———. 2011. "The History of Information Technology." *Annual Review of Information Science and Technology* 45, No. 1: 431–87.

Haigh, Thomas, Mark Priestley, and Crispin Rope. 2014. "Los Alamos Bets on ENIAC: Nuclear Monte Carlo Simulations, 1947-1948." *Annals of the History of Computing, IEEE* 36, No. 3: 42–63.

Hashagen, Ulf. 2013. "The Computation of Nature, Or: Does the Computer Drive Science and Technology?" In *The Nature of Computation. Logic, Algorithms, Applications*, edited by Paola Bonizzoni, Vasco Brattka, and Benedikt Löwe, 7921:263–70. Lecture Notes in Computer Science. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-39053-1_30.

Heide, Lars. 2009. *Punched-Card Systems and the Early Information Explosion, 1880-1945*. Baltimore: Johns Hopkins University Press.
Irani, Keki B., Jie Cheng, Usama M. Fayyad, and Zhaogang Qian. 1993. “Applying Machine Learning to Semiconductor Manufacturing.” *IEEE Expert* 8, No. 1: 41–47.
Jagger, Graham. 2003. “The Making of Logarithm Tables.” In *The History of Mathematical Tables: From Sumer to Spreadsheets*, edited by Martin Campbell-Kelly, Mary Croarken, Raymond Flood, and Eleanor Robson, 49–78. Oxford: Oxford University Press.
Jones, Matthew L. forthcoming. *Reckoning with Matter: Calculating Machines, Innovation, and Thinking about Thinking from Pascal to Babbage*. Chicago: University of Chicago Press.
Kay, Lily E. 2000. *Who Wrote the Book of Life?: A History of the Genetic Code*. Stanford, CA : Stanford University Press.
Kidwell, Peggy A. 1990. “American Scientists and Calculating Machines: From Novelty to Commonplace.” *IEEE Annals of the History of Computing* 12, No. 1: 31–40.
Lenhard, Johannes, Terry Shinn, and Günter Küppers. 2006. “Computer Simulation: Practice, Epistemology, and Social Dynamics.” In *Simulation*, 25:3–22. Sociology of the Sciences Yearbook. Dordrecht: Springer Netherlands. http://dx.doi.org/10.1007/1-4020-5375-4_1.
Leonelli, Sabina. 2014. “What Difference Does Quantity Make? On the Epistemology of Big Data in Biology.” *Big Data & Society* 1, No. 1. doi:10.1177/2053951714534395.
Leonelli, Sabina, and Rachel A. Ankeny. 2012. “Re-Thinking Organisms: The Impact of Databases on Model Organism Biology.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43, No. 1: 29–36. doi:10.1016/j.shpsc.2011.10.003.
Light, Jennifer S. 1999. “When Computers Were Women.” *Technology and Culture* 40, No. 3: 455–83.
Lindgren, Michael. 1990. *Glory and Failure: The Difference Engines of Johann Müller, Charles Babbage and Georg and Edvard Scheutz*. Cambridge, Mass.: MIT Press.
MacKenzie, Donald. 1991. “The Influence of the Los Alamos and Livermore National Laboratories on the Development of Supercomputing.” *Annals of the History of Computing* 13, No. 2: 179–201.
Mahoney, Michael S. 2005. “The Histories of Computing(s).” *Interdisciplinary Science Reviews* 30, No. 2: 119–35. doi:10.1179/030801805X25927.
———. 2011. *Histories of Computing*. Edited by Thomas Haigh. Cambridge, Mass.: Harvard University Press.
Marguin, Jean. 1994. *Histoire des instruments et machines à calculer: Trois siècles de mécanique pensante, 1642-1942*. Paris: Hermann.
McCormick, Bruce H., Thomas A. DeFanti, and Maxine D. Brown. 1987. *Visualization in Scientific Computing*. Vol. 21. Computer Graphics. New York: ACM Press. http://www.sci.utah.edu/vrc2005/McCormick-1987-VSC.pdf.
Metropolis, Nicholas, and Stanislaw Ulam. 1949. “The Monte Carlo Method.” *Journal of the American Statistical Association* 44, No. 247: 335–41.

Mindell, David A. 2002. *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. Baltimore: The Johns Hopkins University Press.

Morgan, Mary S., and Margaret Morrison, eds. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Nolan, Richard L. 2000. "Information Technology Management Since 1960." In *Nation Transformed by Information: How Information Has Shaped the United States from Colonial Times to the Present*, edited by Alfred Dupont Chandler and James W. Cortada, 217–56. New York: Oxford University Press.

November, Joseph Adam. 2012. *Biomedical Computing: Digitizing Life in the United States*. Baltimore, Md.: Johns Hopkins University Press.

Owens, Larry. 1986. "Vannevar Bush and the Differential Analyzer: The Text and Context of an Early Computer." *Technology and Culture* 27, No. 1: 63–95.

Polachek, Harry. 1995. "History of the Journal Mathematical Tables and Other Aids to Computation, 1959-1965." *Annals of the History of Computing* 17, No. 3: 67–74.

Priestley, Mark. 2011. *A Science of Operations*. London: Springer.

Quinlan, J. R. 1979. "Discovering Rules by Induction from Large Collections of Examples." In *Expert Systems in the Micro-Electronic Age*, edited by Donald Michie, 168–201. Edinburgh: Edinburgh University Press.

Rees, Mina. 1950. "The Federal Computing Machine Program." *Science* 112, No. 2921: 731–36.

Schaffer, Simon. 1994. "Babbage's Intelligence: Calculating Engines and the Factory System." *Critical Inquiry* 21: 203–27.

Seidel, Robert W. 1998. "'Crunching Numbers': Computers and Physical Research in the AEC Laboratories." *History and Technology* 15, No. 1-2: 31–68. doi:10.1080/07341519808581940.

Sepkoski, David. 2013. "Towards 'A Natural History of Data': Evolving Practices and Epistemologies of Data in Paleontology, 1800–2000." *Journal of the History of Biology* 46, No. 3: 401–44. doi:10.1007/s10739-012-9336-6.

Simon, Herbert A. 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge, Mass.: MIT Press.

Simpson, George Gaylord. 1962. "Primate Taxonomy and Recent Studies of Nonhuman Primates." *Annals of the New York Academy of Sciences* 102, No. 2: 497–514. doi:10.1111/j.1749-6632.1962.tb13656.x.

Snyder, Samuel S. 1980. "Computer Advances Pioneered by Cryptologic Organizations." *Annals of the History of Computing* 2, No. 1: 60–70.

Stevens, Hallam. 2013. *Life Out of Sequence: A Data-Driven History of Bioinformatics*. Chicago: University of Chicago Press.

Strasser, Bruno J. 2012. "Data-Driven Sciences: From Wonder Cabinets to Electronic Databases." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43, No. 1: 85–87. doi:10.1016/j.shpsc.2011.10.009.

Swade, Doron. 2001. *The Difference Engine: Charles Babbage and the Quest to Build the First Computer*. 1st American Edition. New York: Viking.

Thoren, Victor E. 1988. "Prosthaphaeresis Revisited." *Historia Mathematica* 15, No. 1: 32–39. doi:10.1016/0315-0860(88)90047-X.

Von Neumann, John, and Herman H. Goldstine. 1961. "On the Principles of Large Scale Computing Machines (1946)." In *Collected Works*, by John Von Neumann, 5:1–33. New York: Pergamon Press.

Warwick, Andrew. 1995. "The Laboratory of Theory, Or, What's Exact about the Exact Sciences." In *Values of Precision*, edited by M. Norton Wise, 135–72. Princeton: Princeton Univ. Press.

Winsberg, Eric B. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press, 2010.

Yates, JoAnne. 1993. "Co-Evolution of Information-Processing Technology and Use: Interaction between the Life Insurance and Tabulating Industries." *Business History Review* 67, No. 01: 1–51. doi:10.2307/3117467.

———. 2000. "Business Use of Information and Technology during the Industrial Age." In *Nation Transformed by Information: How Information Has Shaped the United States from Colonial Times to the Present*, edited by Alfred Dupont Chandler and James W. Cortada, 107–36. New York: Oxford University Press.

Yood, Charles N. 2013. *Hybrid Zone: Computers and Science At Argonne National Laboratory, 1946-1992*. Docent Press.

Zhang, Tian, Raghu Ramakrishnan, and Miron Livny. 1996. "BIRCH: An Efficient Data Clustering Method for Very Large Databases." In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, June 4-6, 1996*, edited by H. V. Jagadish and Inderpal Singh Mumick, 103–14. ACM Press.

Biographical Note

Matthew L. Jones teaches at Columbia. A Guggenheim Fellow, he is completing a book on the National Security Agency, and is undertaking a historical and ethnographic account of "big data," its relation to statistics and machine learning, and its growth as a fundamental new form of technical expertise in business, political, and scientific research. His *Reckoning with Matter: Calculating Machines, Innovation, and Thinking About Thinking from Pascal to Babbage* is forthcoming from Chicago.

---

[1] For an overview of the historiography, which has taken a decided turn toward business history, see (Haigh 2011); for sharp historiographical insight on the histories of computing, (Mahoney 2011); for "computing" before the digital computer, with good reference to engineering traditions, see (Akera 2007, chap. 1). The classic study of the early development of the digital computer for scientific applications is (Goldstine 1972). For the spread of information technologies internationally, see (Cortada 2012).

[2] A crucial corrective to simple narratives of computerization is (Agar 2006, 873; compare Hashagen 2013; Mahoney 2005).

[3] Among many studies, see, e.g., (Kay 2000).

[4] For broader concerns about tables, see (Warwick 1995, 317–327).

[5] (Marguin 1994; Aspray 1990; Jones forthcoming)

[6] See (Nolan 2000; Yates 2000; Heide 2009; Warwick 1995; Cortada 2000)

[7] For the UK, see the revisionist account (Agar 2003); for the Soviet Union, see (Crowe and Goodman 1994; Goodman 2003).
[8] For the ENIAC and Monte Carlo, see (Haigh, Priestley, and Rope 2014).
[9] For an international survey, see (Brezinski and Wuytack 2001).
[10] See, however, the fine (Winsberg, 2010). For models in the history of science, see (Morgan and Morrison 1999; Creager, Lunbeck, and Wise 2007).
[11] For histories of data, see, for example, (Leonelli 2014), (Sepkoski 2013; Strasser 2012; Edwards 2010).
[12] The main academic histories of database systems are (Bergin and Haigh 2009; Haigh 2009); more generally, see (Nolan 2000).
[13] See (Burri and Dumit 2008) for visualization studies in STS.