

## 2 Charles Babbage and the Emergence of Automated Reason

Seth Bullock

Charles Babbage (1791–1871) (figure 2.1) is known for his invention of the first automatic computing machinery, the Difference Engine and later the Analytical Engine, thereby prompting some of the first discussions of machine intelligence (Hyman 1982). Babbage's efforts were driven by the need to efficiently generate tables of logarithms—the very word “computer” having originally referred to people employed to calculate the values for such tables laboriously by hand. Recently, however, historians have started to describe the wider historical context within which Babbage was operating, revealing how he, his contemporaries, and their students were influential in altering our conception of the workforce, the workplace, and the economics of industrial production in a Britain increasingly concerned with the automation of labor (Schaffer 1994).

While it was clear that all manner of unskilled manual labour could be achieved by cleverly designed mechanical devices, the potential for the same kind of machinery to replicate mental labor was far more controversial. Were reasoning machines possible? Would they be useful? Even if they were, was their use perhaps less than moral? Babbage's contribution to this debate was typically robust. In demonstrating how computing machinery could take part in (and thereby partially automate) academic debate, he challenged the limits of what could be achieved with mere automata, and stimulated the next generation of “machine analysts” to conceive and design devices capable of moving beyond mere mechanical calculation in an attempt to achieve full-fledged automated reason.

In this chapter, some of the historical research that has focused on Babbage's early machine intelligence and its ramifications will be brought together and summarized. First, Babbage's use of computing within academic research will be presented. The implications of this activity on the wider question of machine intelligence will then be discussed, and the relationship between automation and intelligibility will be explored.



**Figure 2.1**

Charles Babbage in 1847. Source: [http://www.kevyr.net/pioneers/gallery/ns\\_babbage2.htm](http://www.kevyr.net/pioneers/gallery/ns_babbage2.htm) (in public domain).

Intermittently throughout these considerations, connections between the concerns of Babbage and his contemporaries and those of modern artificial intelligence (AI) will be noted. However, examining historical activity through modern lenses risks doing violence to the attitudes and significances of the agents involved and the complex causal relationships between them and their works. In order to guard against the overinterpretation of what is presented here as a “history” of machine intelligence, the paper concludes with some caveats and cautions.

### **The Ninth Bridgewater Treatise**

In 1837, twenty-two years before the publication of Darwin’s *On the Origin of Species* and over a century before the advent of the first modern computer, Babbage published a piece of speculative work as an uninvited Ninth

Bridgewater Treatise (Babbage 1837; see also Babbage 1864, chapter 29, "Miracles," for a rather whimsical account of the model's development). The previous eight works in the series had been sponsored by the will of Francis Henry Egerton, the Earl of Bridgewater and a member of the English clergy. The will's instructions were to make money available to commission and publish an encyclopedia of natural theology describing "the Power, Wisdom, and Goodness of God, as manifested in the Creation" (Brock 1966; Robson 1990; Topham 1992).

In attempting such a description, natural theologians tended to draw attention to states of affairs that were highly unlikely to have come about by chance and could therefore be argued to be the work of a divine hand. For instance, the length of the terrestrial day and seasons seem miraculously suited to the needs and habits of plants, man, and other animals. Natural theologians also sought to reconcile scientific findings with a literal reading of the Old Testament, disputing evidence that suggested an alarmingly ancient earth, or accounting for the existence of dinosaur bones, or promoting evidence for the occurrence of the great flood. However, as Simon Schaffer (1994) points out, natural theology was also "the indispensable medium through which early Victorian savants broadcast their messages" (p. 224).

Babbage's contribution to the Bridgewater series was prompted by what he took to be a personal slight that appeared in the first published and perhaps most popular Bridgewater Treatise. In it, the author, Reverend William Whewell, denied "the mechanical philosophers and mathematicians of recent times any authority with regard to their views of the administration of the universe" (Whewell 1834, p. 334, cited in Schaffer 1994, p. 225). In reply, Babbage demonstrated a role for computing machinery in the attempt to understand the universe and our relationship to it, presenting the first published example of a *simulation model*.

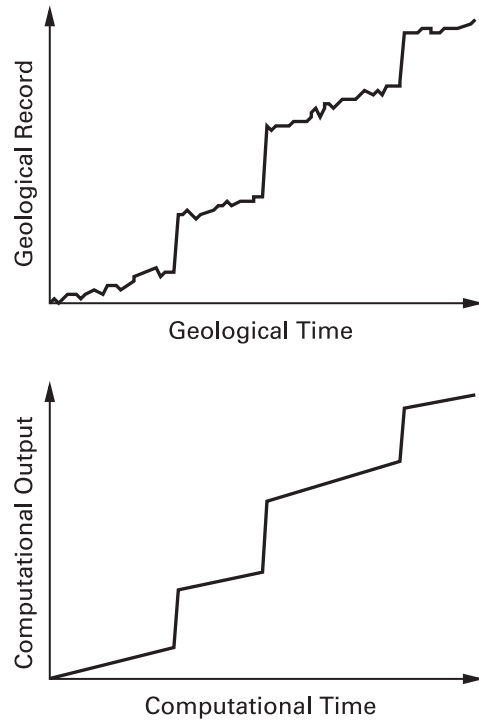
In 1837, Babbage was one of perhaps a handful of scientists capable of carrying out research involving computational modeling. In bringing his computational resources to bear on a live scientific and theological question, he not only rebutted Whewell and advanced claims for his machines as academic as well as industrial tools, but also sparked interest in the extent to which more sophisticated machines might be further involved in full-blown reasoning and argument.

The question that Babbage's model addressed was situated within what was then a controversial debate between what Whewell had dubbed catastrophists and uniformitarians. Prima facie, this dispute was internal to geology, since it concerned the geological record's potential to show evidence

of divine intervention. According to the best field geologists of the day, geological change “seemed to have taken place in giant steps: one geological environment contained a fossil world adapted to it, yet the next stratum showed a different fossil world, adapted to its own environment but not obviously derivable from the previous fossil world” (Cannon 1960, p. 7). Catastrophists argued for an interventionist interpretation of this evidence, taking discontinuities in the record to be indicators of the occurrence of miracles—violations of laws of nature. In contrast, uniformitarians argued that allowing a role for sporadic divine miracles interrupting the action of natural processes was to cast various sorts of aspersions on the Deity, suggesting that His original work was less than perfect, and that He was constantly required to tinker with his Creation in a manner that seemed less than glorious. Moreover, they insisted that a precondition of scientific inquiry was the assumption that the entire geological record must be assumed to be the result of unchanging processes. Miracles would render competing explanations of nature equally valid. No theory could be claimed to be more parsimonious or coherent than a competing theory that invoked necessarily inexplicable exogenous influences. As such, the debate was central to understanding whether and how science and religion might legitimately coexist.

W. Cannon (1960) argues that it is important to recognize that this debate was not a simple confrontation between secular scientists and religious reactionaries that was ultimately “won” by the uniformitarians. Rather, it was an arena within which genuine scientific argument and progress took place. For example, in identifying and articulating the degree to which the natural and physical world fitted each other, both currently and historically, and the startling improbability that brute processes of contingent chance could have brought this about, authors such as Whewell laid a foundation upon which Darwin’s evolutionary theory sat naturally.

Babbage’s response to the catastrophist position that apparent discontinuities were evidence of divine intervention was to construct what can now be recognized as a simple simulation model (see figure 2.2). He proposed that his suitably programmed Difference Engine could be made to output a series of numbers according to some law (for example, the integers, in order, from 0 onward), but then at some predefined point (say 100,000) begin to output a series of numbers according to some different law such as the integers, in order, from 200,000 onward. Although the output of such a Difference Engine (an analogue of the geological record) would feature a discontinuity (in our example the jump from 100,000 to 200,000), the underlying process responsible for this output would have

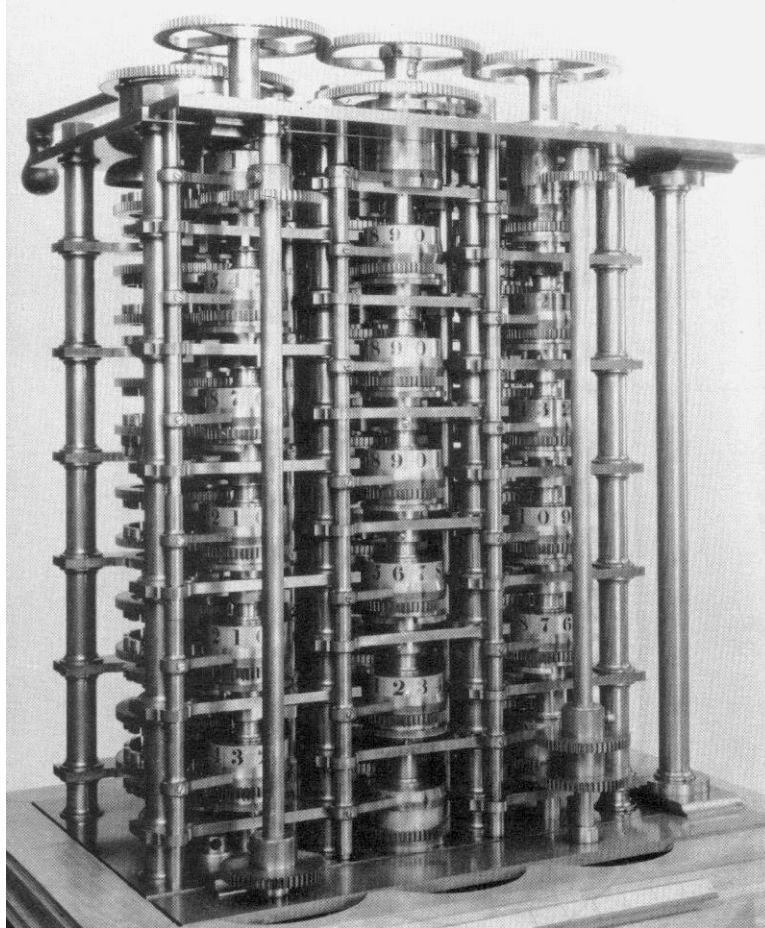


**Figure 2.2**

Babbage's (1836) evolutionary simulation model represented the empirically observed history of geological change as evidenced by the geological record (upper panel) as the output of a computing machine following a program (lower panel). A suitably programmed computing machine could generate sequences of output that exhibited surprising discontinuities without requiring external influence. Hence discontinuities in the actual geological record did not require "catastrophic" divine intervention, but could be the result of "gradualist" processes. Source: Seth Bullock.

remained constant—the general law, or program, that the machine was obeying would not have changed. The discontinuity would have been the result of the naturally unfolding mechanical and computational process. No external tinkering analogous to the intervention of a providential deity would have taken place.

Babbage not only described such a program in print but demonstrated a working portion of his Difference Engine carrying out the calculations described (see figure 2.3). At his Marylebone residence, he surprised a stream of guests drawn from society and academia with machine behavior that suggested a new way of thinking about both automata and miracles.



**Figure 2.3**

Difference Engine. Source: [http://www.kevryr.net/pioneers/gallery/ns\\_babbage5.htm](http://www.kevryr.net/pioneers/gallery/ns_babbage5.htm)  
(in public domain).

D. Swade (1996) describes how Darwin, recently returned from his voyages on the *Beagle*, was urged by Charles Lyell, the leading geologist, to attend one of Babbage's "soirées where he would meet fashionable intelligentsia and, moreover, 'pretty women'" (p. 44). Schaffer (1994) casts Babbage's surprising machine as providing Darwin with "an analogue for the origin of species by natural law without divine intervention" (pp. 225–26).

In trying to show that discontinuities were not necessarily the result of meddling, but could be the natural result of unchanging processes, Babbage cultivated the image of God as a programmer, engineer, or industrialist, capable of setting a process in motion that would accomplish His intentions without His intervening repeatedly. In Victorian Britain, the notion of God as draughtsman of an "automatic" universe, one that would run unassisted, without individual acts of creation, destruction, and so forth, proved attractive. This conception was subsequently reiterated by several other natural philosophers, including Darwin, Lyell, and Robert Chambers, who argued that it implied "a grander view of the Creator—One who operated by general laws" (Young 1985, p. 148). However, here we are less interested in the theological implications of Babbage's work, and more concerned with the manner in which he exploited his computational machinery in order to achieve an academic goal.

Babbage clearly does not attempt to capture the full complexity of natural geology in his machine's behavior. Indeed, the analogy between the Difference Engine's program and the relevant geological processes is a crude one. However, the formal resemblance between the two was sufficient to enable Babbage's point to be made. His computing machine is thus clearly being employed as a model, and a model of a particular kind—an idealized conceptual tool rather than a realistic facsimile intended to "stand in" for the real thing.

Moreover, the model's goal is not to shed light directly on geological discontinuity per se. Its primary function is to force an audience to reflect on their own reasoning processes (and on those of the authors of the preceding eight legitimate Bridgewater Treatises). More specifically, the experiment encourages viewers to (re)consider the grounds upon which one might legitimately identify a miracle, suggesting that a mere inability to understand some phenomenon as resulting from the continuous action of natural law is not sufficient, for the continuous action of some "higher law," one discernible only from a more systemic perspective, could always be responsible. Thus, Babbage's is an "experiment" that brings no new data to light, it generates no geological facts for its audience, but seeks to re-arrange their theoretical commitments.<sup>1</sup>

Babbage approached the task of challenging his audiences' assumptions as a stage magician might have done (Babbage 1837, p. 35):

Now, reader, let me ask how long you will have counted before you are firmly convinced that the engine, supposing its adjustments to remain unaltered, will continue whilst its motion is maintained, to produce the same series of natural numbers? Some minds perhaps are so constituted, that after passing the first hundred terms they will be satisfied that they are acquainted with the law. After seeing five hundred terms, few will doubt; and after the fifty-thousandth term the propensity to believe that the succeeding term will be fifty thousand and one, will be almost irresistible.

Key to his argument was the surprise generated by mechanical discontinuity. That a process unfolding "like clockwork" could nevertheless confound expectation simultaneously challenged the assumed nature of both mechanical and natural processes and the power of rational scientific induction. In this respect, Babbage's argument resonates with some modern treatments of "emergent behavior." Here, nonlinearities in the interactions between a system's components give rise to unexpected (and possibly irreducible, that is, quasi-miraculous) global phenomena, as when, for instance, the presumably simple rules followed by insects generate complex self-regulating nest architectures (Ladley and Bullock 2005), or, indeed, the way in which novel forms can emerge from shape grammars (March 1996a, 1996b). For Babbage, however, any current inability on our part to reconcile some aggregate property with the constitution and organization of the system that gives rise to it is no reason to award the phenomenon special status. His presumption is that for some more sophisticated observer, reconciling the levels of description will be both possible and straightforward, nonlinearity or no nonlinearity.

Additionally, there is a superficial resemblance between the catastrophist debate of the nineteenth century and the more recent dispute over the theory of punctuated equilibria introduced by Niles Eldredge and Stephen Jay Gould (1973). Both arguments revolved around the significance of what appear to be abrupt changes on geological time scales. However, where Babbage's dispute centered on whether change could be explained by one continuously operating process or must involve two different mechanisms—the first being geological processes, the second Divine intervention—Gould and Eldredge did not dispute that a single evolutionary process was at work. They take pains to point out that their theory does not supersede phylogenetic gradualism, but augments it. They wish to account for the two apparent modes of action evidenced by the fossil record—long periods of stasis, short bursts of change—not by invoking

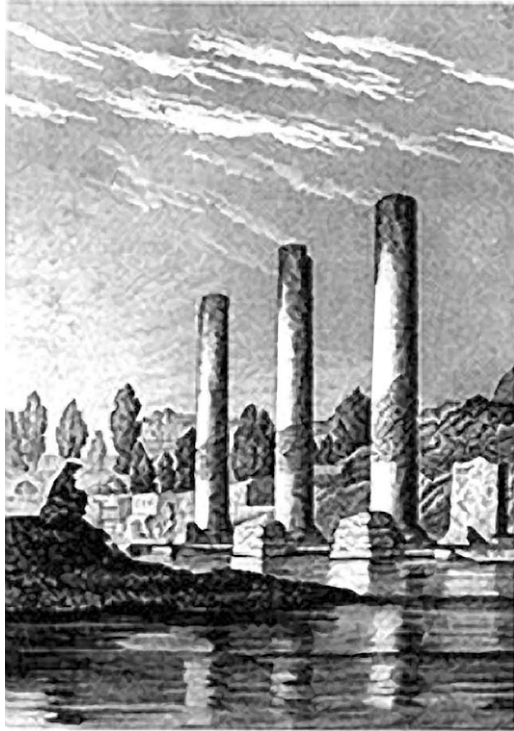


two processes but by explaining the unevenness of evolutionary change. In this respect, the theory that Eldredge and Gould supply attempts to meet a modern challenge: that of explaining nonlinearity, rather than merely accommodating it. Whereas Babbage's aim was merely to demonstrate that a certain kind of nonlinearity was logically possible in the absence of exogenous interference, Gould and Eldredge exemplify the attempt to discover how and why nonlinearities arise from the homogeneous action of low-level entities.

Babbage, too, spent some time developing theories with which he sought to explain how specific examples of geological discontinuity could have arisen as the result of unchanging and continuously acting physical geological processes. One example of apparently rapid geological change that had figured prominently in geological debate since being depicted on the frontispiece of Lyell's *Principles of Geology* (1830) was the appearance of the Temple of Serapis on the edge of the Bay of Baiae in Pozzuoli, Italy (see figure 2.4). The surfaces of the forty-two-foot pillars of the temple are characterized by three regimes. The lower portions of the pillars are smooth, their central portions have been attacked by marine creatures, and above this region the pillars are weathered but otherwise undamaged. These abrupt changes in the character of the surfaces of the pillars were taken by geologists to be evidence that the temple had been partially submerged for a considerable period of time.

For Lyell (1830), an explanation could be found in the considerable seismic activity that had characterized the area historically. It was well known that eruptions could cover land in considerable amounts of volcanic material and that earthquakes could suddenly raise or lower tracts of land. Lyell reasoned that a volcanic eruption could have buried the lower portion of the pillars before an earthquake lowered the land upon which the temple stood into the sea. Thus the lower portion would have been preserved from erosion, while a middle portion would have been subjected to marine perforations and an upper section to the weathering associated with wind and rain.

Recent work by B. P. Dolan (1998) has uncovered the impact that Babbage's own thoughts on the puzzle of the pillars had on this debate. Babbage, while visiting the temple, noted an aspect of the pillars that had hitherto gone undetected: a patch of calciated stone located between the central perforated section and the lower smooth portion. He inferred that this calciation had been caused, over considerable time, by calcium-bearing spring waters that had gradually flooded the temple, as the land upon which it stood sank lower and lower. Eventually this subsidence caused



**Figure 2.4**

The Temple of Serapis. The frontispiece for the first six volumes of Lyell's *Principles of Geology*. By permission of the Syndics of Cambridge University.

the temple pillars to sink below sea level and resulted in the marine erosion evident on the middle portion of the columns.

Thus Babbage's explanation invoked gradual processes of cumulative change, rather than abrupt episodes of discontinuous change, despite the fact that the evidence presented by the pillars is that of sharply separated regimes. Babbage's account of this gradual change relied on the notion that a central, variable source of heat, below the earth's crust, caused expansion and contraction of the land masses above it. This expansion or contraction would lead to subsidence or elevation of the land masses involved. Babbage exploited the power of his new calculating machine in attempting to prove his theory, but not in the form of a simulation model. Instead, he used the engine to calculate tables of values that represented the expansion of granite under various temperature regimes, extrapolated from empirical measurements carried out with the use of furnaces. With

these tables, Babbage could estimate the temperature changes that would have been necessary to cause the effects manifested by the Temple of Serapis (see Dolan 1998, for an extensive account of Babbage's work on this subject).

Here, Babbage is using a computer, and is moving beyond a gradualist account that merely tolerates discontinuities, such as that in his *Bridgewater Treatise*, to one that attempts to explain them. In this case his engine is not being employed as a simulation model but as a prosthetic calculating device. The complex, repetitive computations involved in producing and compiling his tables of thermal expansion figures might normally have been carried out by "computers," people hired to make calculations manually. Babbage was able to replace these error-prone, slow, and costly manual calculations with the action of his mechanical reckoning device.

Like simulation modeling, this use of computers has become widespread across modern academia. Numerical and iterative techniques for calculating, or at least approximating, the results of what would be extremely taxing or tedious problems have become scientific mainstays. However, this kind of automated extrapolation differs significantly from the simulation described above. Just as the word "intelligence" itself can signify, first, the possession or exercise of superior cognitive faculties and, second, the obtainment or delivery of useful information, such as military intelligence, for Babbage, machine intelligence could either refer to some degree of automated reasoning or (less impressively) the "manufacture" of information (Schaffer 1994). While Babbage's model of miracles and his automatic generation of thermal expansion tables were both examples of "mechanized intelligence," they differed significantly in that the first was intended to take part in and thereby partially automate thought processes directed at understanding, whereas the second exemplified his ability to "manufacture numbers" (Babbage 1837, p. 208). This subtle but important difference was not lost upon Babbage's contemporaries, and was central to unfolding discussions and categorizations of mental labor.

### **Automating Reason**

For his contemporaries and their students, the reality of Babbage's machine intelligence and the prospect of further advances brought to the foreground questions concerning the extent to which mental activity could and should be automated. The position that no such activity could be achieved "mechanically" had already been somewhat undermined by the success of unskilled human calculators and computers, who were able to efficiently

generate correct mathematical results while lacking an understanding of the routines that they were executing.

National programs to generate navigational and astronomical tables of logarithmic and trigonometric values (calculated up to twenty-nine decimal places!) would not have been possible in practice without this redistribution of mental effort. Babbage himself was strongly influenced by Baron Gaspard De Prony's work on massive decimal tables in France from 1792, where he had employed a division of mathematical labor apparently inspired by his reading of Adam Smith's *Wealth of Nations* (see Maas 1999, pp. 591–92).

[De Prony] immediately realised the importance of the principle of the division of labour and split up the work into three different levels of task. In the first, "five or six" eminent mathematicians were asked to simplify the mathematical formulae. In the second, a similar group of persons "of considerable acquaintance with mathematics" adapted these formulae so that one could calculate outcomes by simply adding and subtracting numbers. This last task was then executed by some eighty predominantly unskilled individuals. These individuals were referred to as the computers or calculators.

Babbage's Difference Engine was named after this "method of differences," reducing formulae to combinations of addition and subtraction. However, there was a clear gulf separating true thinking from the mindless rote activity of computers, whether human or mechanical. For commentators such as the Italian mathematician and engineer Luigi Federico Menebrea, whose account of a lecture Babbage gave in Turin was translated into English by Ada Lovelace (Lovelace 1843), there appeared little chance that machinery would ever achieve more than the automation of this lowest level of mental activity. In making this judgment, Menebrea "pinpointed the frontiers of the engine's capacities. The machine was able to calculate, but the mechanization of our 'reasoning faculties' was beyond its reach, unless, Menebrea implicitly qualified, the rules of reasoning themselves could be algebraised" (Maas 1999, p. 594–95).

For Menebrea it was apparently clear that such a mental calculus would never be achieved. But within half a century, just such algebras were being successfully constructed by George Boole and John Venn. For some, the potential for mechanizing such schemes seemed to put reasoning machines within reach, but for others, including Venn himself, the objections raised by Menebrea still applied.

Simon Cook (2005) describes how Venn, in his "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings" of 1880, clearly recognized considerable potential for the automation of his logical

formalisms but went on to identify a strictly limited role for such machinery. The nature of the labor involved in logical work, Venn stated (p. 340),

involves four “tolerably distinct steps”: the statement of the data in accurate logical language, the putting of these statements into a form fit for an “engine to work with,” thirdly the combination or further treatment of our premises after such a reduction, and finally interpretation of the results. In Venn’s view only the third of these steps could be aided by an engine.

For Venn, then, computing machinery would only ever be useful for automating the routine process of thoughtlessly combining and processing logical terms that had to be carefully prepared beforehand and the resulting products analyzed afterward.

This account not only echoes De Prony’s division of labor, but, to modern computer scientists, also bears a striking similarity to the theory developed by David Marr (1982) to describe the levels of description involved in cognitive science and artificial intelligence. For Marr, any attempt to build a cognitive system within an information-processing paradigm involves first a statement of the cognitive task in information-processing terms, then the development of an algorithmic representation of the task, before an implementation couched in an appropriate computational language is finally formulated. Venn’s steps also capture this march from formal conception to computational implementation. Rather than stressing the representational form employed at each stage, Venn concentrates on the associated activity, and, perhaps as a result, considers a fourth step not included by Marr: the *interpretation* of the resulting behavior, or output, of the computational process. We will return to the importance of this final step.

Although Venn’s line on automated thought was perhaps the dominant position at that time, for some scholars Babbage’s partially automated argument against miracles had begun to undermine it. Here a computer took part in scientific work not by automating calculation, but in a wholly different way. The engine was not used to compute a result. Rather, the substantive element of Babbage’s model was the manner in which it changed over time. In the scenario that Babbage presented to his audience, his suitably programmed Difference Engine will, in principle, run forever. Its calculation is not intended to produce some end product; rather, the ongoing calculation is itself the object of interest. In employing a machine in this way, as a model and an aid to reasoning, Babbage “dealt a severe blow to the traditional categories of mental philosophy, without positively proving that our higher reasoning faculties could be mechanized” (Maas 1999, p. 593).

Recent historical papers have revealed how the promise of Babbage's simulation model, coupled with the new logics of Boole and Venn, inspired two of the fathers of economic science to design and build automated reasoning machines (Maas 1999; Cook 2005). Unlike Babbage and Lovelace, the names Stanley Jevons (1835–1882) and Alfred Marshall (1842–1924) are not well known to students of computing or artificial intelligence. However, from the 1860s onward, first Jevons and then Marshall brought about a revolution in the way that economies were studied, effectively establishing modern economics. It was economic rather than biological or cognitive drivers that pushed both men to consider the role that machinery might play in automating logical thought processes.

Jevons pursued a mathematical approach to economics, exploring questions of production, currency, supply and demand, and so forth and developing his own system of logic (the "substitution of similars") after studying and extending Boole's logic. His conviction that his system could be automated such that the logical consequences of known states of affairs could be generated efficiently led him to the design of a "logical piano . . . capable of replacing for the most part the action of thought required in the performance of logical deduction" (Jevons 1870, p. 517). But problems persisted, again limiting the extent to which thought could be automated. Jevons's logical extrapolations relied upon the substitution of like terms, such as "London" and "capital of England." The capacity to decide which terms could be validly substituted appeared to resist automation, becoming for Jevons "a dark and inexplicable gift which was starkly to be contrasted with calculative, mechanical rationality" (Maas 1999, p. 613). Jevons's piano, then, would not have inclined Venn to alter his opinion on the limitations of machine logic.

Cook (2005) has recently revealed that Marshall (who, upon Jevons's early death by drowning in 1882, would eventually come to head the marginalist revolution within economics) also considered the question of machine intelligence. In "Ye Machine," the third of four manuscripts thought to have been written in the late 1860s to be presented to Cambridge Grote Club, he described his own version of a machine capable of automatically following the rules of logic. However, in his paper he moves beyond previous proponents of machine intelligence in identifying a mechanism capable of elevating his engine above mere calculation, to the realm of creative reason. Menebrea himself had identified the relevant respect in which these calculating machines were significantly lacking in his original discussion of Babbage's engines. "[They] could not come to any

correct results by 'trial and guess-work', but only by fully written-out procedures" (Maas 1999, p. 593). It was introducing this kind of exploratory behavior that Marshall imagined. What was required were the kinds of surprising mechanical jumps staged by Babbage in his drawing room. Marshall (Cook 2005, p. 343) describes a machine with the ability to process logical rules that,

"like Paley's watch", might make others like itself, thus giving rise to "hereditary and accumulated instincts." Due to accidental circumstances the "descendants," however, would vary slightly, and those most suited to their environment would survive longer: "The principle of natural selection, which involves only purely mechanical agencies, would thus be in full operation."

As such, Marshall had imagined the first example of an explicitly evolutionary algorithm, a machine that would surprise its user by generating and testing new "mutant" algorithmic tendencies. In terms of De Prony's tripartite division of labor, such a machine would transcend the role of mere calculator, taking part in the "adapting of formulae" function heretofore carried out by only a handful of persons "of considerable acquaintance with mathematics." Likewise, Marshall's machine broke free of Venn's restrictions on machine intelligence. In addition to the task of mechanically combining premises according to explicitly stated logics, Marshall's machine takes on the more elevated task of generating new, superior logics and their potentially unexpected results.

Andy Clark (1990) has described the explanatory complications introduced by this move from artificial intelligences that employ explicit, manually derived logic to those reliant on some automatic process of design or adaptation. Although the descent through Marr's "classical cascade" involved in the manual design of intelligent computational systems delivers, as a welcome side effect, an understanding of how the system's behavior derives from its algorithmic properties, no such understanding is guaranteed where this design process is partially automated. For instance, Marr's computational algorithms for machine vision, once constructed, were understood by their designer largely as a result of his gradual progression from computational to algorithmic and implementational representations. The manual design process left him with a grasp of the manner in which his algorithms achieved their performance. By contrast, when one employs artificial neural networks that learn how to behave or evolutionary algorithms that evolve their behavior, a completed working system demands further interpretation—Venn's fourth step—before the way it works can be understood.

The involvement of automatic adaptive processes thus demands a partial inversion of Marr's cascade. In order to understand an adaptive machine intelligence, effort must be expended recovering a higher, algorithmic-level representation of how the system achieves its performance from a working implementation-level representation. The scale and connectivity of the elements making up these kinds of adaptive computational system can make achieving this algorithmic understanding extremely challenging.

For at least one commentator on machine intelligence, it was exactly the suspect intelligibility of automatic machine intelligence that was objectionable. The Rev. William Whewell was a significant Victorian figure, having carved out a role for himself as historian, philosopher, and critic (see figure 2.5). His principal interest was in the scientific method and the role of induction within it. For Whewell, the means with which scientific questions were addressed had a moral dimension. We have already heard how Whewell's dismissal of atheist mathematicians in his *Bridgewater Treatise* seems to have stimulated Babbage's work on simulating miracles (though Whewell was likely to have been targeting the mathematician Pierre-Simon Laplace rather than Babbage). He subsequently made much more explicit



**Figure 2.5**  
The Rev. William Whewell in 1835.



attacks on the use of machinery by scientists—a term he had coined in 1833.

Whewell brutally denied that mechanised analytical calculation was proper to the formation of the academic and clerical elite. In classical geometry “we tread the ground ourselves at every step feeling ourselves firm,” but in machine analysis “we are carried along as in a rail-road carriage, entering it at one station, and coming out of it at another. . . . It is plain that the latter is not a mode of exercising our own locomotive powers. . . . It may be the best way for men of business to travel, but it cannot fitly be made a part of the gymnastics of education. (Schaffer 1994, pp. 224–25)

The first point to note is that Whewell’s objection sidesteps the issues of performance that have occupied us so far. Here, it was irrelevant to Whewell that machine intelligence might generate commercial gain through accurate and efficient calculation or reasoning. A legitimate role within science would be predicated not only on the ability of computing machines to replicate human mental labor but also on their capacity to aid in the revelation of nature’s workings. Such revelation could only be achieved via diligent work. Shortcuts would simply not do. For Whewell it was the journey, not the destination, that was revelatory. Whewell’s objection is mirrored by the assertion sometimes made within artificial intelligence that if complex but inscrutable adaptive algorithms are required in order to obtain excellent performance, it may be necessary to sacrifice a complete understanding of how exactly this performance is achieved—“We are engineers, we just need it to work.” Presumably, Whewell would have considered such an attitude alien to academia.

More prosaically, the manner in which academics increasingly rely upon automatic “smart” algorithms to aid them in their work would have worried Whewell. Machine intelligence as typically imagined within modern AI (for example, the smart robot) may yet be a distant dream, but for Whewell and Babbage, it is already upon us in the automatically executed statistical test, the facts, figures, opinions, and arguments instantaneously harvested from the Internet by search engines, and so forth. Where these shortcuts are employed without understanding, Whewell would argue, academic integrity is compromised.

There are also clear echoes of Whewell’s opinions in the widespread tendency of modern theoreticians to put more faith in manually constructed mathematical models than automated simulation models of the same phenomena. While the use of computers to solve mathematical equations numerically (compare Babbage’s thermal expansion calculations) is typically regarded as unproblematic, there is a sense that the complexity—the

impenetrability—of simulation models can undermine their utility as scientific tools (Grimm 1999; Di Paolo et al. 2000).

However, it is in Marshall's imagined evolving machine intelligence that the apotheosis of Whewell's concerns can be found. In the terms of Whewell's metaphor, not only would Marshall be artificially transported from problem to solution by such a machine, but he would be ferried through deep, dark, unmapped tunnels in the process. At least the rail tracks leading from one station to another along which Whewell's imagined locomotive must move had been laid by hand in a process involving much planning and toil. By contrast, Marshall's machine was free to travel where it pleased, arriving at a solution via any route possible. While the astonishing jumps in the behavior of Babbage's machine were not surprising to Babbage himself, even the programmer of Marshall's machine would be faced with a significant task in attempting to complete Venn's "interpretation" of its behavior.

### Conclusion

This chapter has sought to highlight activities relevant to the prehistory of artificial intelligence that have otherwise been somewhat neglected within computer science. In gathering together and presenting the examples of early machine intelligence created by Babbage, Jevons, and Marshall, along with contemporaneous reflections on these machines and their potential, the chapter relies heavily on secondary sources from within a history of science literature that should be of growing importance to computer science. Although this paper attempts to identify a small number of issues that link contemporary AI with the work of Babbage and his contemporaries, it is by no means a piece of historical research and the author is no historian. Despite this, in arranging this material here on the page, there is a risk that it could be taken as such.

Babbage's life and work have already been the repeated subject of Whigish reinterpretation—the tendency to see history as a steady linear progression (see Hyman 1990 for a discussion). In simplifying or ignoring the motivations of our protagonists and the relationships between them, there is scope here, too, for conveying the impression of an artificially neat causal chain of action and reaction linking Babbage, Whewell, Jevons, Marshall, and others in a consensual march toward machine intelligence driven by the same questions and attitudes that drive modern artificial intelligence. Such an impression would, of course, be far from the truth. The degree to which each of these thinkers engaged with questions of machine intelli-

gence varied wildly: for one it was the life's work; for another, a brief interest. And even with respect to the output of each individual, the elements highlighted here range from significant signature works to obscure footnotes or passing comments. It will be left to historians of science to provide an accurate account of the significances of the activities presented here. This chapter merely seeks to draw some attention to them.

Given the sophistication already evident in the philosophies associated with machine intelligence in the nineteenth century, it is perhaps surprising that a full-fledged philosophy of technology, rather than science, has only recently begun to emerge (Ihde 2004). In the absence of such a discipline, artificial intelligence and cognitive philosophy, especially that influenced by Heideggerian themes, have played a key role in extending our understanding of the role that technology has in influencing the way we think (see, for example, Dreyfus 2001). If we are to cope with the rapidly expanding societal role of computers in, for instance, complex systems modeling, adaptive technologies, and the Internet, we must gain a firmer grasp of the epistemic properties of the engines that occupied Babbage and his contemporaries.

Unlike an instrument, that might simply be a pencil, engines embody highly differentiated engineering knowledge and skill. They may be described as "epistemic" because they are crucially generative in the practice of making scientific knowledge. . . . Their epistemic quality lies in the way they focus activities, channel research, pose and help solve questions, and generate both objects of knowledge and strategies for knowing them. (Carroll-Burke 2001, p. 602)

### Acknowledgments

This chapter owes a significant debt to the painstaking historical research of Simon Schaffer, B. P. Dolan, S. Cook, H. Maas, and, less recently, W. Cannon.

### Note

1. See Bullock (2000) and Di Paolo, Noble, and Bullock (2000) for more discussion of Babbage's simulation model and simulation models in general.

### References

Babbage, Charles. 1837. *Ninth Bridgewater Treatise: A Fragment*. 2nd edition. London: John Murray.

- Brock, W. H. 1966. "The Selection of the Authors of the Bridgewater Treatises." *Notes and Records of the Royal Society of London* 21: 162–79.
- Bullock, Seth. 2000. "What Can We Learn from the First Evolutionary Simulation Model?" In *Artificial Life VII: Proceedings of the Seventh International Conference On Artificial Life*, edited by M. A. Bedau, J. S. McCaskill, N. H. Packard, and S. Rasmussen. Cambridge, Mass.: MIT Press.
- Cannon, W. 1960. "The Problem of Miracles in the 1830s." *Victorian Studies* 4: 4–32.
- Carroll-Burke, P. 2001. "Tools, Instruments and Engines: Getting a Handle on the Specificity of Engine Science." *Social Studies of Science* 31, no. 4: 593–625.
- Clark, A. 1990. "Connectionism, Competence and Explanation." In *The Philosophy of Artificial Intelligence*, edited by Margaret A. Boden. Oxford: Oxford University Press.
- Cook, S. 2005. "Minds, Machines and Economic Agents: Cambridge Receptions of Boole and Babbage." *Studies in the History and Philosophy of Science* 36: 331–50.
- Darwin, Charles. 1859. *On the Origin of Species*. London: John Murray.
- Di Paolo, E. A., J. Noble, and Seth Bullock. 2000. "Simulation Models as Opaque Thought Experiments." In *Artificial Life VII: Proceedings of the Seventh International Conference On Artificial Life*, edited by M. A. Bedau, J. S. McCaskill, N. Packard, and S. Rasmussen. Cambridge, Mass.: MIT Press.
- Dolan, B. P. 1998. "Representing Novelty: Charles Babbage, Charles Lyell, and Experiments in Early Victorian Geology." *History of Science* 113, no. 3: 299–327.
- Dreyfus, H. 2001. *On the Internet*. London: Routledge.
- Eldredge, N., and Steven Jay Gould. 1973. "Punctuated Equilibria: An Alternative to Phyletic Gradualism." In *Models in Paleobiology*, edited by T. J. M. Schopf. San Francisco: Freeman, Cooper.
- Grimm, V. 1999. "Ten Years of Individual-Based Modelling in Ecology: What We Have Learned and What Could We Learn in the Future?" *Ecological Modelling* 115: 129–48.
- Hyman, R. A. 1982. *Charles Babbage: Pioneer of the Computer*. Princeton: Princeton University Press.
- . 1990. "Whiggism in the History of Science and the Study of the Life and Work of Charles Babbage." *IEEE Annals of the History of Computing* 12, no. 1: 62–67.
- Ihde, D. 2004. "Has the Philosophy of Technology Arrived?" *Philosophy of Science* 71: 117–31.
- Jevons, W. S. 1870. "On the Mechanical Performance of Logical Inference." *Philosophical Transactions of the Royal Society* 160: 497–518.

- Ladley, D., and Seth Bullock. 2005. "The Role of Logistic Constraints on Termite Construction of Chambers and Tunnels." *Journal of Theoretical Biology* 234: 551–64.
- Lovelace, Ada. 1843. "Notes on L. Menabrea's 'Sketch of the Analytical Engine invented by Charles Babbage, Esq.'" *Taylor's Scientific Memoirs*. Volume 3. London: J. E. & R. Taylor.
- Lyell, Charles. 1830/1970. *Principles of Geology*. London: John Murray; reprint, London: Lubrecht & Cramer.
- Maas, H. 1999. "Mechanical Rationality: Jevons and the Making of Economic Man." *Studies in the History and Philosophy of Science* 30, no. 4: 587–619.
- March, L. 1996a. "Babbage's Miraculous Computation Revisited." *Environment and Planning B: Planning & Design* 23(3): 369–76.
- . 1996b. "Rulebound Unruliness." *Environment and Planning B: Planning & Design* 23: 391–99.
- Marr, D. 1982. *Vision*. San Francisco: Freeman.
- Robson, J. M. 1990. "The Fiat and the Finger of God: The Bridgewater Treatises." In *Victorian Crisis in Faith: Essays on Continuity and Change in 19th Century Religious Belief*, edited by R. J. Helmstadter and B. Lightman. Basingstoke, U.K.: Macmillan.
- Schaffer, S. 1994. "Babbage's Intelligence: Calculating Engines and the Factory System." *Critical Inquiry* 21(1): 203–27.
- Swade, D. 1996. "'It Will Not Slice a Pineapple': Babbage, Miracles and Machines." In *Cultural Babbage: Technology, Time and Invention*, edited by F. Spufford and J. Uglow. London: Faber & Faber.
- Topham, J. 1992. "Science and Popular Education in the 1830s: The Role of the Bridgewater Treatises." *British Journal for the History of Science* 25: 397–430.
- Whewell, W. 1834. *Astronomy and General Physics Considered with Reference to Natural Theology*. London: Pickering.
- Young, R. M. 1985. *Darwin's Metaphor: Nature's Place in Victorian Culture*. Cambridge: Cambridge University Press.

