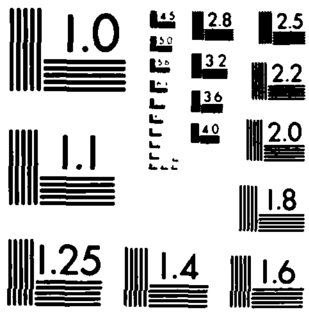
AD-A126 438 STUDYING SCIENTIFIC DISCOVERY BY COMPUTER SIMULATION (U) CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF PSYCHOLOGY G L BRADSHAW ET AL. 30 MAR 83 1/1

UNCLASSIFIED CMU-CIP-WP-444 N00014-82-K-0168 F/G 9/2 NL

END
DATE
FILMED
4 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

CIP Working Paper No. 444

Studying Scientific Discovery by Computer Simulation[1]

Gary L. Bradshaw
Pat Langley
Herbert A. Simon
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

[1] This research was supported by Contract N00014-82-K-0168 from the Office of Naval Research, and by Research Grant MH-07722 from the National Institutes of Mental Health. G. L. Bradshaw and H. A. Simon are associated with the CMU Department of Psychology, while P. Langley is affiliated with The Robotics Institute.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report No. 2	2. GOVT ACCESSION NO. AD-A126 938	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Studying Scientific Discovery by Computer Simulation		5. TYPE OF REPORT & PERIOD COVERED Interim Report 10/82-3/83
		6. PERFORMING ORG. REPORT NUMBER CIP No. 444
7. AUTHOR(s) Gary L. Bradshaw Patrick W. Langley Herbert A. Simon		8. CONTRACT OR GRANT NUMBER(s) N00014-82-K-0168
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Robotics Institute Carnegie-Mellon University Pittsburgh, Pennsylvania 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 049-514
11. CONTROLLING OFFICE NAME AND ADDRESS Information Sciences Division Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE 30 March, 1983
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Submitted to Science.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

scientific discovery	intrinsic properties
physical laws	extensive terms
data-driven heuristics	intensive terms
theory-driven heuristics	conservation laws

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Scientific discovery is a complex process, but despite its complexity, it can be analyzed using two basic ideas from the theory of human problem solving - the notion of search through a problem space, and the notion of heuristics for directing that search down profitable paths. In this paper we examine Black's discovery of the specific heat law in the light of these concepts, and we consider how BACON, an empirical discovery program, employs heuristics for finding numeric relations, positing intrinsic

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. properties, and expecting symmetrical forms to rediscover this law. Finally, we examine the role of theory in scientific discovery, and describe a different approach to inducing Black's law that relies on the distinction between extensive and intensive terms, and on the notion of conservation.

Studying Scientific Discovery by Computer Simulation

Gary L. Bradshaw, Pat Langley, and Herbert A. Simon
Carnegie-Mellon University

Scientists, artists, and other creative people, being generously endowed with curiosity, often direct that curiosity to their own mental processes. Perhaps more than curiosity is involved. Perhaps they believe or hope that a deeper understanding of their processes of discovery will allow them to be more creative. However that may be, a number of distinguished scientists have undertaken to speak and write on the topic of creativity, among the best known accounts being those of Poincare[2] and Hadamard.[3] They have provided us with a rich and consistent description of the phenomena that are visible during the work of discovery, identifying and labeling such characteristic aspects and phases as "preparation," "incubation," "intuition," and the "aha!" experience.

Psychologists and sociologists, too, have directed their attention to invention, discovery, and the creative processes. They have interviewed and questioned scientists and artists, examined histories, biographies, and autobiographies, and even observed creativity (at least at modest levels) in the laboratory. Their phenomenological accounts agree very well with those provided by the scientists and engineers themselves. In addition, they have learned something about the interaction of scientific discovery with its social environment – both within the scientific discipline and in the larger society.

None of this work, however, gives us much of an explanation of how a discovery actually comes about, and in particular, how it can be brought about by an information processing system having the characteristics of the human brain. Words like "incubation" and "intuition" do not tell us what is going on; they merely emphasize that much of what is going on occurs without detailed conscious awareness. Instead of explaining the process to us, these terms explain that little of the process is open to conscious reports.

Computer Simulation Methodology

During the past quarter century considerable success has been attained in constructing computer programs that simulate in some detail the processes that people use to solve relatively difficult problems (e.g., making a chess move, understanding a verbal description of a puzzle). This progress has suggested the extension of computer simulation techniques to studying the processes of creative thinking, including scientific discovery. In fact, as early as 1958 the hypothesis had been proposed that the underlying processes of creative thinking were essentially the same as those that were needed to account for more ordinary garden varieties of problem solving.[4] What distinguished creative thinking from other forms of thought, it was argued, was *not qualitatively different processes*, but, instead, the poorly structured character – the "vagueness" – of the problems attacked, and the creative person's persistence, over long periods of time, in attacking them.

Rather than speculate about the hypothesis that creativity is problem solving writ large, we can begin to subject this hypothesis to empirical tests. One way to do this is through computer simulation.

---

[2] "Mathematical creation," p. 387 in *The Foundations of Science*, N.Y.: The Science Press, 1913; reprinted in Volume 4 of James R. Newman, *The World of Mathematics*, N.Y.: Simon & Schuster, 1956.

[3] *The Psychology of Invention in the Mathematical Field*, Princeton: Princeton U. Press, 1945

[4] A. Newell, J. C. Shaw, and H. A. Simon, "The processes of creative thinking," reprinted in H. A. Simon, *Models of Thought*, pp. 144-174 (1979). See also, H. A. Simon, "Scientific discovery and the psychology of problem solving," in R. Colodny (ed.), *Mind and Cosmos* (1966), pp.22-40.

First, we specify the problem solving processes we think are required for discovery, defining them with such specificity that we can write a computer program to execute them. Second, we confront the computer program with discovery problems that have presented themselves to scientists,and we observe the program's behavior – not only whether or not it succeeds in solving the problem, but also whether it goes through the same steps as human solvers, or quite different steps.

To carry out this comparison between the computer program and human behavior, we must obtain data about the latter. In the work to be described here, these data take the form mainly of historical accounts of actual scientific discoveries of first importance that were made by physicists and chemists in the 17th, 18th, and 19th Centuries. These accounts are not wholly adequate for testing the computer program as a theory of discovery, for they provide only coarse-grained narratives of what took place. We will use them here mainly to identify the initial conditions, what was already known at the time the discovery was made and how the problem was conceptualized by the discoverer and his contemporaries. We will not attempt detailed, step-by-step comparisons of the discovery processes, both because the empirical data do not exist for that and because our own theory of discovery is also coarse grained, and purports to describe the processes only qualitatively. The most important information we will be able to glean from our simulations is whether the processes that we postulate, and that are incorporated in the computer program, are *sufficient* to account for the new discovery on the basis of the data that were available to, and used by, the discoverer.

The simulation approach is best illustrated by a concrete example. Before we introduce the example, however, one more caveat is in order. Scientific discovery involves not just a single process, but a whole array of processes, including gathering data, finding parsimonious descriptions of the data, and formulating and testing explanatory theories. In this paper, we will be concerned with just part of the whole process – principally with *data-driven discovery*, that is, discovery that starts with given data, and derives new descriptive or explanatory laws from them. Of course, the generation of data, and even the invention of instruments to produce new kinds of data, are also important aspects of scientific discovery. And in many cases, pre-existing theory as well as empirical data steer the course of discovery. We will have comments to make on these other components of the discovery process, but our attention will be directed initially to the path from data to laws. We will show that a large amount of important scientific work falls in this category.

As an example, we turn now to the discovery of Black's Law: the law describing the equilibrium temperature of a mixture of two liquids.

Black's Discovery

About 1760 – the exact date is not known – Joseph Black, a Scottish chemistry professor, made the first of the several important discoveries that have preserved his name.[5] Using data reported in a standard chemistry textbook of his time (Boerhaave's) and obtained from an experiment performed at Boerhaave's request by Fahrenheit, Black reinterpreted the data to formulate what we now know as Black's law for the temperature of mixtures, or Black's law of heat. Since Fahrenheit reported his invention of the mercury thermometer in 1724 and died in 1736, the data that Black used must have been some 25 to 35 years old.

The experiment that produced the data was very simple: mixing two substances while measuring their initial and final temperatures. It is, of course, an everyday experience that the temperature of hot water can be moderated by mixing it with cold. What Fahrenheit and other early

[5]Many of the source documents relating to this and most of the other scientific discoveries discussed in this paper can be found in H. M. Leicester and H. S. Klockstein, *A Source Book in Chemistry*, N.Y.: McGraw-Hill, 1952; and W. F. Magie, *A Source Book in Physics*, N.Y.: McGraw-Hill, 1935. On Black's Law see Magie, pp. 134-9.

investigators, armed with the newly-invented thermometer, showed was that if equal volumes of water at different temperatures are mixed, the mixture will have a temperature that is the arithmetic average of the initial temperatures. Other experiments showed that if the original amounts are unequal, the average has to be weighted by those initial amounts.

It had also been discovered (and was reported by Boerhaave) that if two objects of *different* substance with different initial temperatures are mixed, they will come to a final temperature that is some kind of average of their original temperatures. However, there was great confusion about the nature of that average, it being commonly maintained, in the face of contrary empirical evidence, that the average depended on the masses or volumes involved, but not on the species of substances that were mixed. What contributed especially to the confusion was the fact that a heavy substance, like mercury, appeared to influence the final temperature less than an equal volume of a lighter substance, like water. Apparently this result was so counterintuitive as to impede clear understanding.

Joseph Black cleared up the confusion, and formulated the law that is named after him, simply by attributing a new property to the unit volume (or mass)[6] of each particular substance, its *specific heat*. The specific heat multiplied by the object's volume (or mass) gave its *heat capacity* The final temperature of a mixture could be calculated by averaging the initial temperatures, using the heat capacities as weights.

Different substances had different specific heats, which could be determined from the mixing experiment. No independent, converging procedures were available for estimating or predicting specific heats until Dulong and Petit, in 1819, showed they were approximately inversely proportional to atomic weights. And no theoretical basis for the empirical values was forthcoming until very much later than that. Black was in most respects a thoroughgoing empiricist. Nevertheless, he was willing to introduce these theoretical constructs, the specific heats, solely in order to preserve the property of conservation of total heat in mixture experiments using several substances.

Inductive Derivation of Laws by BACON

It would seem, at first blush, that Black's accomplishment was a pure act of Baconian induction. Confronted with a set of data on the equilibrium temperatures that resulted from mixing quantities of mercury and water, he introduced new constants, the specific heats. He then found that the final temperature could be predicted as a function of the initial temperatures, volumes and specific heats, the latter a property of each substance. But how could this induction be achieved? The failure of Fahrenheit, Boerhaave, and others to arrive at it over the thirty years after the experiments were performed suggests that it was not a trivial matter. What kind of induction process could overcome the difficulties?

The technique of computer simulation of cognitive processes allows us to answer at least part of this question. In particular, a program known as BACON, when provided with data like those that Black had, soon discovers that they can be summarized parsimoniously by the formula we know as Black's law. In the course of making the discovery, BACON also invents the concept that Black called "specific heat."

BACON was not initially constructed with Black's law in mind (the initial versions were debugged mainly with data for Kepler's Third Law, Ohm's law, and Snell's law of refraction), but is a general data-driven system for discovering scientific laws.[7] The program's initial inputs are data that

[6] Black saw that it did not matter which of the two measures of amount was used.

[7] A description of the basic BACON program will be found in H. A. Simon, P. W. Langley, and G. L. Bradshaw, "Scientific discovery as problem solving," *Synthese*, 47:1-27 (1981).

may include both numerical and nominal values; its outputs are empirical laws that summarize those data. A number of versions of BACON have existed in the course of its development, later versions generally having richer sets of heuristics than the earlier ones. Our account here will be based mainly on BACON.4 and BACON.5.

BACON begins to examine the independent variables given it, one at a time, while holding the others constant, with the aim of discovering a functional relation between independent and dependent variables. If a relation is found, additional variables are introduced, one at a time, and BACON attempts to generalize the function it has found to include each new variable. This process is continued until an invariant relation is found that holds for the entire body of data, or until BACON runs out of time. BACON has a modest capacity to ignore noise in the data.

BACON's search for invariants is selective, under the guidance of a few powerful heuristics. It detects constancies in data. It looks for linear relations between variables. When one variable varies monotonically with respect to another, BACON examines the product or ratio of the two as a possible invariant. For example if the current in a circuit varies inversely with the resistance (i.e., in the absence of internal resistance), BACON computes their product, finding it to be constant. If the product or ratio is not constant, it is still introduced as a new variable (a theoretical term), and treated exactly like the original observational variables.

When BACON discovers that a relation among certain variables depends on a nominal variable, it associates a new numerical variable (*intrinsic property*) with the nominal variable. For example, if the ratio of current to resistance in a circuit depends on which of several batteries is inserted in the circuit, BACON will associate with the nominal variable **battery** a new numerical variable, which we would call the **voltage** of the battery, and which is measured by the current/resistance ratio. Once defined, the intrinsic properties are treated like the other variables of the system, and can be retrieved later when useful for the analysis. It is this ability to introduce new intrinsic properties in association with nominal variables that enables BACON to create the variable *specific heat* to account for the different effects of different substances upon the equilibrium temperature in the mixture experiments.

In some situations, two objects (e.g., two batteries, two resistances) have exactly the same set of properties, although with different values. In this case, BACON.5 assumes that the law it is seeking will be symmetrical with respect to the objects, although it cannot predict the actual form of the law. It orders the data so that variables attached to one of the objects are varied first, the others being held constant. Once a constant function has been found incorporating all the variables associated with the one object, BACON will assume that an analogous function should be defined for the other object. In this way, the symmetrical laws so common in physics are discovered with great saving in search. This assumption of symmetry can be very helpful in discovering Black's law of specific heat, although it is not essential for that purpose.

Although BACON's search is primarily driven by the data, its use of the hypothesis of symmetry shows that it can also respond to theoretical considerations – as presumably human scientists can. We shall presently outline a version of BACON that also can entertain hypotheses of conservation (in this case, conservation of heat) as a heuristic for facilitating its discovery of laws.

BACON's Derivation of Black's Law

Black's Law can be stated quite simply. If one brings two containers of liquid at different temperatures into contact, their temperatures will gradually equalize. The quantitative relation is:

$$c_1m_1t_1 + c_2m_2t_2 = (c_1m_1 + c_2m_2)f$$

where t_1 and t_2 are the initial temperatures, f is the final temperature, m_1 and m_2 are the masses of the two liquids, and c_1 and c_2 are the specific heats associated with the two species of liquids. If we call the coefficients of the two temperatures on the left side of the equation A and B, respectively, then we can see that the final temperature is the weighted average of the temperatures of the components, with weights of A/(A + B) and B/(A + B), respectively. That is, $f = (At_1 + Bt_2) / (A + B)$. We recall that the coefficients A and B are the heat capacities of the two substances, while At_1 and Bt_2 are their total heats, and $(A + B)f$ is the total heat of the mixture. Hence, the equation can be interpreted as saying that total heat is conserved, where the heat capacity of the mixture is the sum of the heat capacities of the components.

Now we can sketch out an inductive path that will lead from the experimental data to Black's law. For simplicity of exposition, we will not follow exactly the path used by any single version of BACON, but will describe one that does not differ in essential respects from those that were actually tested successfully.

At the outset, we employ identical volumes of the same substance (say water), at different temperatures. We mix them and let the mixture come to equilibrium. In the first experiment, we vary the temperature of the first component, holding all other variables constant. We discover that the equilibrium temperature is a linear function of the independent variable, with a slope of 1/2. In the second experiment, we vary the second temperature, using symmetry between the two volumes to *conjecture* that the equilibrium temperature will also vary linearly with this new independent variable, with a slope of 1/2. The data confirm this relation.

In the third experiment, we vary the mass of the first component in order to determine how that mass enters the function determining the equilibrium temperature. Having accomplished that, we now conjecture that a symmetrical function will describe the joint effects of both masses. The data confirm this conjecture.

In the fifth experiment, we change the composition of the first component (for example, substitute mercury for water). BACON finds that a new coefficient must be introduced into the equation, whose value changes with change in the composition of the component. In the sixth experiment, we change the composition of the second component, and, by symmetry, introduce another new coefficient into the equation (using the same specific heats that were inferred before). The equation, thus modified, again fits the data. The *coefficients introduced in the fifth and sixth* experiments are, of course, the quantities that Black christened specific heats.[8]

Examples of Data-Driven Induction

It should not be thought that Black's law is an isolated example of data-driven induction. In the history of modern science, it has not been at all uncommon for the discovery of laws of first importance to precede, sometimes by many years, the discovery of theoretical conceptions that could account for the lawfulness of the data. We will cite a few examples from the 17th, 18th, and 19th centuries, and from the fields of astronomy, physics, and chemistry.

Kepler's Third Law, expressing the periods of revolution of the planets about the sun as the 3/2 power of their distances from it, was a product of pure Baconian induction. It was a hundred years later that Newton produced the gravitational theory from which Kepler's law could be deduced. No theory in Kepler's day constrained the functional form of the relation.

Ohm's Law is another clearcut case where data preceded any substantial theoretical framework,[9] although the analogy of hydraulic flow may have been helpful. The invention of the

---

[8]We have described the last stages of the inductive process only in the roughest terms. Parenthetically, it should be mentioned here that the derivation of Black's law by BACON.4, reported in Bradshaw, Langley, and Simon, "BACON.4: The discovery of intrinsic properties," *Proceedings of the Third National Conference of the Canadian Society for Computational Studies of Intelligence* (1980), pp. 19-25 is incomplete, but the derivation by BACON.5, in Langley, Bradshaw, and Simon, "The discovery of conservation laws," *Proceedings of the 7th International Joint Conference on Artificial Intelligence* (1981), pp. 121-126, is correct.

[9]See Magie, pp. 465-472

Voltaic cell around 1800 transformed electricity from a transient phenomena – a spark produced by a Leyden jar – into a continuous one – a current in a circuit. Later, Seebeck's thermopile permitted reasonably constant voltages to be maintained. The subsequent invention by Ampere, about 1820, of an instrument for measuring the strength of the current provided a quantifiable observable for experiments with circuits. The second observable (the independent variable) that Ohm selected was the length of the resistance wire in the circuit. The discovery of Ohm's law in 1826 became then a straightforward task of inducing the relation between current and resistance.

Balmer was a teacher of geometry in a Swiss gymnasium who attended a physics lecture in 1884 at which the four then known lines of the hydrogen spectrum were mentioned. Struck by the apparent regularity of the wave lengths, and using only the data for the first four lines in the spectrum, Balmer induced the formula associated with his name and predicted correctly the wave lengths of additional lines.[10] No theoretical basis for his formula existed until Bohr created one some thirty years after its discovery. Balmer's feat was a matter of pure induction, using both arithmetic and geometric heuristics.

Prout, on the basis of the data on atomic weights available around 1815, noticed that most of the weights were nearly integral multiples of the atomic weight of hydrogen.[11] There were notable exceptions (e.g., chlorine), but these did not prevent Prout from making the generalization. The explanation both of Prout's law and its exceptions had to await the invention of the mass spectrograph in 1919.

This by no means exhausts the list of scientific laws that were induced from data before any theory was available to discover the regularities. To the previous examples, we could add Gregor Mendel's laws of inheritance, the law of Gay-Lussac for gaseous reactions, the law of Dulong and Petit, the derivation of atomic weights by Avogadro and Cannizzaro, and the construction of the periodic table by Mendeleev. Experiments with BACON have demonstrated paths along which most of these discoveries (excluding those of Mendel and Mendeleev, which we have not simulated) could have been made.

The Role of Theory in Law Induction

BACON's derivation of Black's Law, just outlined, is truly data driven. It was achieved simply by carrying out a succession of controlled experiments, each introducing a new independent variable that had previously been held constant. Out of the experimental data emerged not only the law, but a new theoretical concept, the specific heat. Theoretical presuppositions played only a minor role in the discovery – BACON assumed that the two substances being mixed would play symmetrical roles in the final equilibrium. This assumption reduced the amount of search required to find the final form of the law, but other versions of BACON have found the same law, more tediously, without this assumption.

However, data-driven induction is not the only route to discovery. The opposite extreme, of course, is the case where there already exists a theory so complete and powerful that the laws, the parsimonious description of the data, can be deduced from it without any induction at all. The idea of a critical experiment depends on theory being able to predict the data from new experiments. Lying somewhere between pure induction and pure deduction are situations where more or less vague and incomplete theoretical constructs already exist that can guide and channel somewhat the course of the induction, but that fall considerably short of permitting a priori deduction of the actual laws.

A look at the circumstances surrounding the discovery of Black's law provides evidence that

---

[10] See Magie, pp. 360-365

[11] See Leicester and Klickstein, pp. 275-279

the actual discovery was, indeed, aided and guided by some conceptual framework, and was not as completely data driven as the pathway taken by BACON. In his posthumous textbook, Black summarizes briefly the evidence for the law that "all bodies communicating freely with each other, and exposed to no inequality of external action, acquire the same temperature as indicated by a thermometer." He then observes that the nature of this equilibrium was not well understood; in particular, that it was sometimes supposed to imply that there was an equal amount of heat per unit volume of each body. "But," he says, "*this is taking a very hasty view of the subject. It is confounding the quantity of heat in different bodies with its general strength or intensity, though it is plain that these are two different things, and should always be distinguished, when we are thinking of the distribution of heat.*"

While the equilibrium experiments measured only one physical variable, the temperature, Black here contrasts "quantity of heat" with "general strength or intensity," and attributes the difficulties of his predecessors to their failure to make a sharp distinction between these and to [illegible]ntify temperature with the latter rather than the former. The idea of quantity of heat predat[illegible] [illegible]lack, playing a central role in the caloric theory. Although the very terms "hot" and "cold" a[illegible] [illegible]ed in everyday language refer to heat intensity, it is not clear that this idea had been equal[illegible] [illegible]early conceptualized before the time of Black. Let us see, however, how he might have used a ve[illegible] [illegible]eral conceptual framework, involving the relation between extensive and intensive quantities, [illegible]fy these notions.

We are familiar with extensive quantities like mass and volume. These quantities are additive – the volume of the combination of two objects is the sum of the volumes of the individual objects, and similarly with mass. The density, the mass per unit volume, behaves differently. The density of the combined object is a *weighted* average of the densities of the components, their respective volumes constituting the weights.

But there is an easier way to think of the density in terms of conservation principles. When objects are combined, mass and volume are conserved.[12] Density is simply the ratio of mass to volume. Hence, to compute the density of the combination, simply divide the total mass by the total volume. The notion of weighted averages need not enter if the extensive, conserved, quantities are given primary position in the representation, and intensive quantities only a secondary position.

With this general framework at hand, one can tackle the problem of the heat experiments. Black may well have approached the experimental data with the idea already somewhat in mind that quantity of heat is an extensive magnitude, and moreover, one that is conserved. He could also have had the idea that large quantities of substances would contain, other things being equal, greater quantities of heat than small quantities of substances. Hence, one could define an intensive quantity by dividing the total heat by a substance's capacity for heat (presumably, its mass or volume). Black identified this ratio with temperature.

Black's predecessors were probably not wholly confused on this point. They were clearly prepared to admit that large quantities of water would have a greater influence on the equilibrium temperature in a mixture experiment than small quantities. Their problem came from identifying "large quantity," in terms of volume or weight, with "heat capacity." Before the fact, both volume and mass were plausible candidates for the measure of heat capacity, volume because it defined the space within which the heat, or caloric, was held, mass because it defined the "quantity of matter" of the heat-bearing substance. While we do not have any historical evidence on this point, we would suppose that volume would appeal as the plausible variable to scientists committed to a caloric theory, while mass would seem more plausible to someone holding a "motion" theory of heat.

---

[12] Of course this is not always true; whether it is in any given case is a matter of physics and chemistry. But early in life we encounter many situations where such conservation holds, and we build up a representation for thinking about that broad class of situations.

What Black perceived, and his predecessors evidently did not, was that neither of these extensive measures of heat capacity was consistent with the data from the experiments. He also saw that by introducing a new property, with a different and initially wholly empirical value for each substance, he could express total heat as the product of the extensive variable, heat capacity, and the intensive variable, temperature. Heat capacity, in turn, could be expressed as a product of the traditionally accepted variables, mass or volume, by a new intensive quantity, specific heat. In mixture experiments with a single substance, the terms in specific heat simply cancel out, leaving the classical results in which volume or mass can serve as the measure of capacity. When different subjects are mixed, the new coefficient takes care of the inequality of their effects on the final temperature.

We would not like to insist on the detailed accuracy of this account of how Black arrived at the concept of specific heat, for his posthumous textbook enjoys the clarity of hindsight, and does not necessarily represent his state of mind at the time of the discovery. What is important for present purposes is that this somewhat hypothetical sequence shows how pre-theoretical ideas could guide the interpretation of new data. In the next section, we will outline a version of BACON that is capable of making use of conceptualizations of this kind to aid its data analysis.

Explaining Concept-Guided Induction

We have seen that volume, mass, and density constitute a system of related quantities, two extensive and one intensive. When two objects, each possessing a definite volume, mass, and density, are combined, the resulting object has mass and volume each equal to the sum of the masses and volumes, respectively, of the components, while the density of the combined object is a weighted average of the densities of the components. The weights are the volumes of the components.

The same relations hold among any triad of extensive and intensive variables. where the third is the ratio of the first two. If the extensive quantities are conserved (i.e., if they are additive under composition), then the intensive quantity will be a weighted average of the component quantities, the weights being the denominators of the ratios defining the intensive quantity.

Suppose, now, that we give BACON the information that C and H are extensive and conservative quantities, while $T = H / C$. We also provide data, as in the experiments previously described, for the *initial and equilibrium values of T under different conditions*. BACON can then employ the relations:

$$H_1 + H_2 = C_1T_1 + C_2T_2 = C_fT_f = H_f$$
$$C_f = C_1 + C_2$$

Fitting the data to these relations, it will be easy to show that the C's vary proportionately with the masses of their respective objects, but that they are different for equal masses of different substances. If mass is also accepted as an extensive. conservative measure, then the variations in C can be "explained" by introducing $c = C / M$, where c, the specific heat, is a property of the substance employed in the experiment.[13] Using these definitional identities, Black's Law can now be stated in precisely the form that was displayed earlier.

The conservation assumptions are therefore very powerful, allowing the system to predict the form of the law, and requiring the data only to show the need for specific heats and to estimate their values. Inductive search is essentially eliminated. In fact, so "obvious" does the derivation of the law become that it is difficult to understand why highly intelligent and creative scientists in the 18th century had such great trouble in arriving at it.

There were two paths, then, a purely inductive one and a deductive one, that could have been

[13] The system just sketched has not yet been implemented, but appears to be within the current state of the art.

used to discover Black's law. The historical evidence does not make very clear their relative roles in the discovery, and hindsight, far from clarifying the matter, makes it even harder to understand. We will leave the historical example here, for our main motive in introducing it was to illustrate a few of the routes that scientific discovery can take, and the kinds of mechanisms that must be available if these routes are to be traversed.

Conclusion

The experiments with the BACON program show how scientific discoveries of first magnitude can be arrived at using only data-driven induction processes, unaided by theoretical concepts. In fact, the inductive processes, in the course of discovering laws, themselves create new theoretical terms, including new intrinsic properties. The BACON experiments help to explain, thereby, how some of the important laws of modern science could antecede by decades or even generations the invention of the theoretical structures that subsequently rationalized them and took them out of the category of brute empirical generalizations.

On the other hand, the experiments also show that introducing "pre-theoretical" constructs, like symmetry and conservation, may reduce significantly the amount of search required to detect empirical regularities in data. In fact, if sufficiently strong hypotheses are available, the roles of theory and data may be reversed, so that laws are now deduced directly from theoretical assumptions and subsequently tested by data.

Our brief analysis of the history of the discovery of Black's Law suggests that it was neither wholly data-driven, nor wholly deduced from theoretical assumptions, but that both induction from experimental observations and deduction from notions of the conservation of extensive properties played a role in the discovery.

More generally, the experiments with BACON provide a set of hypotheses about how scientific discovery comes about. The processes that BACON uses are much the same as those that have emerged from other research on human problem solving, and which are usually described as processes for selective heuristic search through spaces of possible solutions. The less "blind" the search – i.e., the more pre-existing theoretical knowledge is available to guide it and turn it from unprofitable directions – the more readily and directly it discovers the regularities hidden in empirical data.

DATE
FILMED
8