Inductionist Exercises

Scientific Discovery. Computational Explorations of the Creative Process. PAT LANGLEY, HERBERT A. SIMON, GARY L. BRADSHAW, and JAN M. ZYTKOW. MIT Press, Cambridge, MA, 1987. viii, 358 pp., illus. $25; paper, $9.95.

The belief that truth in empirical science is most efficiently discovered through the use of some canons of "inductive method" has had many defenders. That it has lately become a minority view among philosophers of science may be a consequence of the revolutions in physics early in this century. These events demonstrated the vulnerability of carefully developed and firmly held theories and thereby cast doubt on the value of any philosophical reconstruction of the methodology employed in their construction. For whatever reason, recent discussions of scientific methodology have tended to distinguish sharply between a "context of discovery" and a "context of verification," focusing most of their attention on the latter. Some have argued that attempts to reconstruct "inductive methods" are necessarily doomed to failure. The only task left for scientific methodology is clarifying the canons of "acceptance" of scientific claims—whatever their source might be.

This book provides a sustained argument for reintroducing problems of inductive methodology into mainstream philosophy of science. It argues that both descriptive and normative accounts of inductive methodology are feasible goals, that discovery and verification cannot be sharply distinguished, and that reconstructions of inductive methods applied to historical examples might be useful tools in the history of science. It does this mainly by providing concrete examples of inductive methods embodied in computer programs, demonstrating that they are successful in "rediscovering" certain scientific truths from data, and finally arguing that these successes are models of historically significant scientific discovery.

The inductive methods described are intended to discover several different kinds of scientific truths. But all are informed by the authors' view that "scientific discovery is problem solving writ large" (p. 6) together with a clearly defined conception of what is involved in "human problem solving." This conception comes in part from earlier work of Simon and others in cognitive psychology described in chapter 1. Problem solving is conceived as a search through a space of possible solutions, with the acceptability of each solution evaluated in turn. This search is not random but is guided by "inductive heuristics," which are more effective than random search but fall short of guaranteeing a solution. This conception applies most naturally to "well-structured" problems—problems whose possible solutions may be exhaustively described and for which there is an effective means of testing a solution's correctness. The authors admit that not all scientific discovery can be viewed as the solution of well-structured problems, suggesting that the applicability of the model may be limited to what Kuhn has called "normal science" (p. 16).

The first application of these ideas (chapters 3 through 5) is a sequence of increasingly more sophisticated heuristics aimed at discovering laws—polynomial equations—describing numerical data. Kepler's third law—$D^3/P^2 = K$—is such a law describing numerical data about the distance (D) of a planet from the sun and its period (P). These heuristics are "data-driven" in the sense that the form of the laws is not constrained by any theory and "weak" in the sense that they are independent of any knowledge of the subject matter to which the data pertain.

The simplest of the authors' heuristics (BACON.1) operates on data with only two numerical quantities. One quantity is designated as the "independent variable"; the remaining is the "dependent variable." Different values of the independent variable are generated and corresponding values of the dependent variable in the data are recorded. The heuristic then looks for simple relations—that is, constant or linear—among the values of the independent and dependent variables. If there is a simple relation, the heuristic returns this relation as the desired law. Failing to find a simple relation, it defines a new dependent variable as either the product or quotient of the independent and dependent variable. It then recurses to check for simple relations between the independent variable and the newly defined dependent variable. This process is iterated until a simple relation is discovered or (presumably) some arbitrarily chosen upper limit of iterations is reached. Intuitive sketches are given of its operation in "rediscovering" Boyle's law, Galileo's law of free fall, and simple forms of Ohm's law and Kepler's third law.

Extending these heuristics to deal with more than two numerical quantities (BACON.3) yields rediscoveries of the ideal gas law, Coulomb's law, and more complex versions of Ohm's law and Kepler's third law. Further extensions (BACON.4) with the capacity to define "theoretical concepts"—intrinsic properties of individuals like specific heat and inertial mass—yield rediscoveries of Snell's law, Black's law, and collision mechanical momentum conservation, among others. An additional procedure for noting common divisors enables the discovery of Dalton's law of simple multiple proportions from data about combining weights in chemical reactions. Still further extensions (BACON.5) relating to symmetry and conservation considerations lead to more efficient rediscoveries of the same laws.

Recognizing limits inherent in heuristics dealing only with numerical data (p. 192), the authors turn their attention to heuristics aimed at discovering qualitative laws and structural models (chapters 6 through 8). These are focused on examples from 18th- and 19th-century chemistry, though further potential applications are sketched (p. 271 ff.).

GLAUBER (chapter 6) is a system designed to produce concepts (taxonomic clusters of individuals) and laws involving these concepts from relational data about individuals. Its data types are exemplified by relational data about inputs and outputs of chemical reactions of the form

reacts (inputs {I1, I2}, outputs {I3})
tastes {I1, sour},

where the I1's refer to individual chemical substances like sodium chloride. The output is classes of individuals like "acid," "base," and "salt" together with laws like

For all X in acid, Y in base, there exists Z in salt such that
(reacts inputs {X, Y}, outputs {Z}).
For all X in acid (tastes {X, sour}).

STAHL (chapter 7) takes the same type of relational data about "reacts" and produces output about "components" of substances of the form

components_of (S1, {C1, C2}).

Roughly, STAHL identifies some substances as "elements" and discovers the elementary composition of other substances. DALTON (chapter 8) attributes structural formulas to substances (for example H_2O to water) on the basis of information about their components in the form generated by STAHL. It is "theory-driven" in the sense that it "employs a representation of chemical reactions and substances that embodies the [atomic] hypothesis." That is, it embodies Dalton's atomic theory simply in the way it formulates the problem of assigning structural formulas.

A brief but informative discussion relates these heuristics to the artificial intelligence literature. GLAUBER is observed to differ from other conceptual clustering heuristics in its use of relational input data. DALTON is compared with DENDRAL—a heuristic for identifying topological structure of organic molecules on the basis of mass spectrographic data—and it is noted that DALTON, though theory-driven, is "weaker" than DENDRAL in that it embodies less domain-specific knowledge.

Despite the substantial ingenuity exhibited in these inductive heuristics and the impressive list of successful "rediscoveries," it remains to be shown that these successes model genuinely significant examples of scientific discovery. The authors argue for this claim in a variety of ways throughout the book, but most explicitly in chapters 2 and 9 through 11. In summary, their arguments are these. First, discovery of the type they have modeled is found to be an essential *part* of clearly significant "discovery events," which also involve other kinds of discovery—for example, Planck's discovery of the law of blackbody radiation was a part of the discovery of quantum mechanics. Second, some other important types of discovery—problem identification, choice of relevant data and their representation, conceptual innovation—are amenable to formulation as "solution of well-structured problems." The limits to which the authors wish to push this second argument are not completely clear, and it is questions about these limits that will provoke controversy.

Some questions arise even about the BACON heuristics. These are capable of law discovery with a given vocabulary and of "conceptual innovation" in discovering intrinsic, theoretical concepts. But both these capacities appear somewhat limited. In the former case the heuristics apparently lack the capacity to discover the appropriate scope for the laws they discover. Given data about two-particle collisions, and having discovered inertial mass and momentum conservation, could such heuristics then discover that kinetic energy conservation as well holds for a subclass of this data? In the latter case the heuristics appear limited in two ways. First, theoretical concepts can be identified only in experimental situations where their values are uniquely determined by the data. Inertial mass cannot be identified from data about collisions involving several particles, in which the velocity data do not suffice to determine it uniquely. Second, it is not apparent how these heuristics might be extended to work in situations where more than one theoretical concept must be identified and employed in the same law. How could they identify both mass and force and discover Newton's second law from kinematic data? This would apparently involve first "postulating" the second law and then discovering some special force law that, together with the second law, described the kinematic data.

There is a sense in which discoveries of interrelated theoretical concepts like those of Newtonian dynamics are discoveries of "novel representations." More obvious examples are the discovery of geocentric epicycle-deferent models for celestial motion and Dalton's atomic hypothesis. The authors minimize the extent to which the discovery of "wholly novel representations" is associated with scientific discovery (p. 333) while also suggesting that a "problem solution" approach to discovery might be extended to discover new representations (p. 326). The examples mentioned may be just the kind of examples of "revolutionary science" that the authors exclude from the scope of their methods. Yet how and where the limits on conceptual innovation by their methods are to be drawn remains an open question.

Questions like these about the "scope and limits" of the account of scientific discovery offered in this book should be the focus of detailed and rigorous analysis among philosophers and historians of science as well as cognitive psychologists and members of the artificial intelligence community. Whatever the verdict turns out to be, the authors are to be credited with renewing the discussion of a traditional question at a higher level of precision and sophistication.

JOSEPH D. SNEED
*Department of Humanities and Social Science,
Colorado School of Mines, Golden, CO 80401*

An Effort at R&D

RCA and the VideoDisc. The Business of Research. MARGARET B. W. GRAHAM. Cambridge University Press, New York, 1986. xiv, 258 pp., illus. $19.95. Studies in Economic History and Policy: The United States in the Twentieth Century.

In this book Graham provides a fascinating account of an unsuccessful technological innovation, the video disc developed by RCA. She describes RCA as a firm that was once the acknowledged leader in television technology, one that took advantage of opportunities, was committed to research, and, unlike most firms, maintained its employment of scientists during the Great Depression. The firm had difficulty repeating its earlier successes, however, and recognized as early as the 1960s that it needed another major product to supplement television sets. Graham describes how RCA decided upon the video recorder: such a device would utilize both its television and its recording technologies and as a leading-edge technology would offer the possibility of cost reductions and proprietary advantages that would yield longer term profits.

RCA decided to work on the video disc in preference to magnetic tape, which was favored by the Consumer Electronics Division early in the project, for what were then valid reasons, including a desire to keep the cost and price of its video recorder low to attract a mass market. Accordingly, RCA pursued the strategy of perfecting a relatively inexpensive machine with few features. The video disc also seemed attractive because the sale of discs would yield continuing profits as with razor blades for razor manufacturers. Attractive programming then would be especially important in selling the system, a factor that eventually caused difficulties because, whereas magnetic tape technology permits the user to make recordings from broadcasts, the video disc is limited to playing previously recorded programs.

Graham details carefully and well the scientific and commercial development of the video disc. She provides examples of the role that competitive factors and internal firm dynamics—especially relations between the laboratory and the operating division—played in the process. The bureaucracy of a large firm, where repeated demonstrations and reports and evaluations were necessary, seemed to slow down the innovative process. Graham also describes in detail how RCA's writeoff of its computer operation caused problems in the development of the disc. By the time RCA finally introduced the video disc player in 1981 the video cassette recorder had caught on and was viewed as superior to video disc recording. Graham suggests that if RCA had been quicker the video disc might have been more successful. She also gives an interesting account of RCA's consideration of adopting a "fast-second" role.

The development of the video disc confirms that the process of innovation can be long, expensive, and uncertain. This reviewer would have appreciated more analysis of why RCA neglected the possibility of Japanese competition and of why it did not anticipate the possibility of substantial cost reduction in magnetic tape technology or the rental of tapes. One also wonders how close to cost RCA priced its machines and whether it could have priced them lower and made more on the discs.

In any case, Graham has given us an excellent and provocative analysis of the process of innovation within a large firm. It