

- Crawford E, Shinn T, Sörlin S 1993 *Denationalizing Science. The Contexts of International Scientific Practice*. Kluwer, Dordrecht, The Netherlands
- Hoskin K W 1993 Education and the genesis of disciplinaryity: The unexpected reversal. In: Messer-Davidow E, Shumway D R, Sylvan D J (eds.) *Knowledges: Historical and Critical Studies in Disciplinaryity*. University Press of Virginia, Charlottesville, VA, pp. 271–305
- Kuhn T S 1970 *The Structure of Scientific Revolutions*, 2nd edn. University of Chicago Press, Chicago
- Luhmann N 1990 *Die Wissenschaft der Gesellschaft*. Suhrkamp, Frankfurt am Main, Germany
- Marrou H I 1934 'Doctrina' et 'Disciplina' dans la langue des pères de l'église. *Archivus Latinitatis Medii Aevi* 9: 5–25
- Messer-Davidow E, Shumway D R, Sylvan D J (eds.) 1993 *Knowledges: Historical and Critical Studies in Disciplinaryity*. University Press of Virginia, Charlottesville, VA
- Ong W J 1958 *Ramus, Method, and the Decay of Dialogue: From the Art of Discourse to the Art of Reason*. Harvard University Press, Cambridge, MA
- Parsons T 1977 *Social Systems and the Evolution of Action Theory*. Free Press, New York
- Stichweh R 1984 *Zur Entstehung des Modernen Systems Wissenschaftlicher Disziplinen—Physik in Deutschland 1740–1890*. Suhrkamp, Frankfurt am Main, Germany
- Stichweh R 1990 Self-organization and autopoiesis in the development of modern science. In: Krohn W, Küppers G, Nowotny H (eds.) *Selforganization—Portrait of a Scientific Revolution. Sociology of the Sciences*, Vol. XIV. Kluwer Academic Publishers, Boston, pp. 195–207
- Stichweh R 1992 The sociology of scientific disciplines: On the genesis and stability of the disciplinary structure of modern science. *Science in Context* 5: 3–15
- Stichweh R 1996 Science in the system of world society. *Social Science Information* 35: 327–40
- Swoboda W W 1979 Disciplines and interdisciplinarity: A historical perspective. In: Kockelmans J J (ed.) *Interdisciplinarity and Higher Education*. University Park, London, pp. 49–92
- Westman R S 1980 The astronomers's role in the sixteenth century: A preliminary study. *History of Science* 18: 105–47

R. Stichweh

Scientific Discovery, Computational Models of

Scientific discovery is the process by which novel, empirically valid, general, and rational knowledge about phenomena is created. It is, arguably, the pinnacle of human creative endeavors. Many academic and popular accounts of great discoveries surround the process with mystery, ascribing them to a combination of serendipity and the special talents of geniuses. Work in Artificial Intelligence on computational models of scientific reasoning since the 1970s shows that such accounts of the process of science are largely mythical. Computational models of scientific discovery are computer programs that make discover-

ies in particular scientific domains. Many of these systems model discoveries from the history of science or simulate the behavior of participants solving scientific problems in the psychology laboratory. Other systems attempt to make genuinely novel discoveries in particular scientific domains. Some have produced new findings of sufficient worth that the discoveries have been published in mainstream scientific journals. The success of these models provides some insights into the nature of human cognitive processes in scientific discovery and addresses some interesting issues about the nature of scientific discovery itself (see *Scientific Reasoning and Discovery, Cognitive Psychology of*).

1. Computational Models of Scientific Discovery

Most computational models of discovery can be conceptualized as performing a recursive search of a space of possible states, or expressions, defined by the representation of the problem. Procedures are used to search the space of legal states by manipulating the expressions and using tests of when the goal or subgoals have been met. To manage the search, which is typically subject to potential combinatorial explosion, heuristics are used to guide the selection of appropriate operators. This is essentially an application of the theory of human problem solving as heuristic search within a symbol processing system (Newell and Simon 1972).

For example, consider BACON (Langley et al. 1987) an early discovery program which finds algebraic formulas as parsimonious descriptions of quantitative data. States in the problem search space of BACON include simple algebraic formulas; such as P/D or P^2/D , where, for instance, P is the period of revolution of planets around the sun and D is their distance from the sun. Tests in BACON attempt to find how closely potential expressions match the given quantitative data. Given quantitative data for the planets of the solar system, one step in BACON's discovery path finds that neither P^2/D nor P/D are constant and that the first expression is monotonically increasing with respect to the second. Given this relation between the expressions BACON applies its INCREASING operator to give the product of the terms, i.e., P^3/D^2 . This time the test of whether the expression is constant, within a given margin of error, is true. $P^3/D^2 = \text{constant}$ is one of Kepler's planetary motion laws. For more complex cases with larger numbers of variables, BACON uses discovery heuristics based on notions of symmetry and the conservation of higher order terms to pare down the search space. The heuristics use the underlying regularities within the domain to obviate the need to explore parts of the search space that are structurally similar to previously explored states.

Following such an approach, computational models have been developed to perform tasks spanning a full

spectrum of theoretical activities including the formation of taxonomies, discovering qualitative and quantitative laws, creation of structural models and the development of process models (Langley et al 1987, Shrager and Langley 1990, Cheng 1992). The range of scientific domains covered is also impressive, ranging from physics and astronomy, to chemistry and metallurgy, to biology, medicine, and genetics. Some systems have produced findings that are sufficiently novel to be worthy of publication in major journals of the relevant discipline (Valdes-Perez 1995).

2. Scope of the Models

Computational models of scientific discovery have almost exclusively addressed theory formation tasks. However, the modeling of experiments has not been completely neglected as models have been built that design experiments, to a limited extent, by using procedures to specify what properties should be manipulated and measured in an experiment and the range of magnitudes over which the properties should be varied (Kulkarni and Simon 1988, Cheng 1992). For these systems, the actual experimental results are either provided by the user of the system or generated by a simulated experiment in the software. Discovery systems have also been directly connected to a robot which manipulates a simple experimental setup so that data collected from the instruments can be fed to the system directly, so eliminating any human intervention (Huang and Zytkow 1997). Nevertheless, few systems have simulated or supported substantial experimental activities, such as observing or creating new phenomena, designing experiments, inventing new experimental apparatus, developing new experimental paradigms, establishing the reliability of experiments, or turning raw data into evidence. This perhaps reflects a fundamental difference between the theoretical and experimental sides of science. While both clearly involve abstract conceptual entities, experimentation is also grounded in the construction and manipulation of physical apparatus, which involves a mixture of sophisticated perceptual abilities and motor skills. Developing models of discovery that include such capabilities would necessarily require other areas within AI beyond problem solving, such as image processing and robotics.

The majority of discovery systems model a single theory formation task. The predominance of such systems might be taken as the basis for a general criticism of computational scientific discovery. The models are typically poor imitations of the diversity of activities in which human scientists are engaged and, perhaps, it is from this variety that scientific creativity arises. Researchers in this area counter such arguments by claiming that the success of such single task systems is a manifestation of the underlying nature of the

process of discovery, that it is composed of subprocesses or tasks that are relatively autonomous. More complex activity can be modeled by assembling systems that perform one task into a larger system, with the inputs to a particular component subsystem being the outputs of other systems. The handful of models that do perform multiple tasks demonstrate the plausibility of this claim (e.g., Kulkarni and Simon 1988, Cheng 1992). The organization of knowledge structures and procedures in those systems exploits the hierarchical decomposition of the overall process into tasks and subtasks.

This in turn raises the general question about the number and variety of different tasks that constitute scientific discovery and the nature of their interactions. What distinct search spaces are involved and how is information shared among them? Computational models of scientific discovery provide some insight into this issue. At a general level, many models can be characterized in terms of two spaces, one for potential hypotheses and the other a space of instances or sets of data (Simon and Lea 1974). Scientific discovery is then viewed as the search of each space mutually constrained by the search of the other. Inferring a hypothesis dictates the form of the data needed to test the hypothesis, while the data itself will determine whether the hypothesis is correct and suggest in what ways it should be amended (Kulkarni and Simon 1988). This is an image of scientific discovery that places equal importance on theory and experiment, portraying the overall process as a dynamic interaction between both components. This approach is applicable both to disciplines in which individual scientists do the theorizing and experimenting and to disciplines in which these activities are distributed among different individuals or research groups. The search of the theoretical and experimental spaces can be further decomposed into additional subspaces; for example, Cheng (1992) suggests three subspaces for hypotheses, models, and cases under the theory component, and spaces of experimental classes, setups, and tests under the experimental component.

3. Developing Computational Models of Discovery

One major advantage of building computational models over other approaches to the study of scientific discovery is the precision that is imposed by writing a running computer program. Ambiguities and inconsistencies in the concepts used to describe discovery processes become apparent when attempting to encode them in a programming language. Another advantage of modeling is the ability to investigate alternative methods or hypothetical situations. Different versions of a system may be constructed embodying, say, competing representations to investigate the difficulty

of making the discovery with the alternatives. The same system can be run with different sets of data, for example, to explore whether a discovery could have been made had there been less data, or had different data been available.

Many stages are involved in the development of the models, including: formulation of the problem, engineering appropriate problem representations, selecting and organizing data, design and redesign of the algorithm, actual invocation of the algorithm, and filtering and interpretation of the findings of the system (Langley 1998). Considering the nature and relative importance of these activities in the development of systems provides further insight into the nature of scientific discovery. In particular, the design of the representation appears to be especially critical to the success of the systems. This implies that generally in scientific discovery finding an effective representation may be fundamental to the making of discoveries. This issue has been directly addressed by computational models that contrast the efficacy of different representations for modeling the same historical episode (Cheng 1996). Consistent with work in cognitive science, diagrammatic representations may in some cases be preferable to informationally equivalent propositional representations. Although computational models argue against any special abilities of great scientists beyond the scope of conventional theories of problem solving, the models suggest that the ability of some scientists to modify or create new representations may be an explanation, at least in part, of why they were the ones to succeed.

4. Conclusions and Future Directions

Given the extent of the development work necessary on a discovery system, it seems appropriate to attribute discoveries as much to the developer as to the system itself, although without the system many of the novel discoveries would not have been possible. This does not imply that machine discovery is impossible, but that care must be taken in delimiting the capabilities of discovery systems. Further, the ability of the KEKEDA system (Kulkarni and Simon 1988) to change its goals to investigate any surprising phenomenon it discovers suggests that systems can be developed that would filter and interpret the output of existing systems, by constraining the search of the space defined by the outputs of those systems using metrics based on notions of novelty. Developing such a system, or other systems that find problems or that select appropriate representations, will require the system to possess a substantially more extensive knowledge of the target domain. Such knowledge based systems are costly and time consuming to build, so it appears that the future of discovery systems will be more as collaborative support systems for domain

scientists rather than fully autonomous systems (Valdes-Perez 1995). Such systems will exploit the respective strengths of domain experts and the computational power of the models to compensate for each others' limitations.

See also: Artificial Intelligence: Connectionist and Symbolic Approaches; Artificial Intelligence in Cognitive Science; Artificial Intelligence: Search; Deductive Reasoning Systems; Discovery Learning, Cognitive Psychology of; Intelligence: History of the Concept; Problem Solving and Reasoning, Psychology of; Problem Solving: Deduction, Induction, and Analogical Reasoning; Scientific Reasoning and Discovery, Cognitive Psychology of

Bibliography

- Cheng P C-H 1992 Approaches, models and issues in computational scientific discovery. In: Keane M T, Gilhooly K (eds.) *Advances in the Psychology of Thinking*. Harvester-Wheatsheaf, Hemel Hempstead, UK, pp. 203–236
- Cheng P C-H 1996 Scientific discovery with law encoding diagrams. *Creativity Research Journal* 9(2&3): 145–162
- Huang K-M, Zytkow J 1997 Discovering empirical equations from robot-collected data. In: Ras Z, Skowron A (eds.) *Foundations of Intelligent Systems*. Springer, Berlin
- Kulkarni D, Simon H A 1988 The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, 12: 139–75
- Langley P 1998 The computer-aided discovery of scientific knowledge. In: *Proceedings of the First International Conference on Discovery Science*. Springer, Berlin
- Langley P, Simon H A, Bradshaw G L, Zytkow J M 1987 *Scientific Discovery: Computation Explorations of the Creative Processes*. MIT Press, Cambridge, MA
- Newell A, Simon H A 1972 *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ
- Shrager J, Langley P (eds.) 1990 *Computational Models of Scientific Discovery and Theory Formation*. Morgan Kaufmann, San Mateo, CA
- Simon H A, Lea G 1974 Problem solving and rule induction: a unified view. In: Gregg L W (ed.) *Knowledge and Cognition*. Lawrence Erlbaum, Potomac, MD, pp. 105–127
- Valdes-Perez R E 1995 Some recent human/computer discoveries in science and what accounts for them. *AI Magazine* 16(3): 37–44

P. C.-H. Cheng

Scientific Evidence: Legal Aspects

Expertise, scientific and otherwise, has been part of the legal landscape for centuries (Hand 1901). Over the last decades of the twentieth century the role of scientific evidence in the law has expanded rapidly, in both regulatory settings and in litigation. Statutes and

13733