

Recent Work in Computational Scientific Discovery

Lindley Darden (darden@umiacs.umd.edu)

Committee on the History and Philosophy of Science, Department of Philosophy
University of Maryland, College Park, MD 20742 USA

Abstract

This paper reviews work in computational scientific discovery. After a brief discussion of its history, the focus will be on work since 1990. The second half of the paper discusses the author's use of three methods for studying reasoning strategies in scientific change: historical-philosophical vs. live-in-the-lab vs. computational, pointing out advantages and disadvantages of the computational method.

Introduction

There are a number of approaches to the study of reasoning in scientific discovery. In addition to computational approaches, work continues in cognitive science (e.g., Schunn & Dunbar, 1996), in laboratory studies (e.g., Darden & Cook, 1994; Dunbar, 1995) and in philosophy of science (e.g., Bechtel & Richardson, 1993; Darden, 1991; Kleiner, 1993; Nersessian, 1992; Nickles, 1994; Schaffner, 1993; Spirtes, Glymour & Scheines, 1993). Unfortunately, of the over 200 papers and abstracts submitted for the Philosophy of Science Association meeting in 1996, none were on the topic of reasoning in scientific discovery (Darden, Ed., 1996; 1997). Most philosophers of science do not view discovery as a central topic in the field, despite continuing work by those of us called "friends of discovery" (Nickles, Ed. 1980). It is encouraging that the Cognitive Science Society is sponsoring this Symposium on Scientific Discovery.

This paper will briefly review the history of computational scientific discovery that uses methods from artificial intelligence. (Non-cognitive, non-AI computational work is outside the scope of this paper.) The first part of the paper will concentrate on the work since 1990 (Shrager & Langley, Eds.). The extensive reference list provides a guide for further reading. The second half of the paper will compare three methods used in my own work on reasoning strategies in scientific change. Finally, I will point out advantages and disadvantages of the computational approach from my perspective as a philosopher of science.

Pioneering Work

The study of computational scientific discovery emerged from the view that science is a problem solving activity, that heuristics for problem solving can be applied to the study of scientific discovery in either historical or contemporary cases, and that methods in artificial intelligence provide techniques for building computational systems. Pioneers in this work are Bruce Buchanan (e.g., 1982) and Herbert Simon (e.g., 1977). Buchanan was trained as a philosopher of science at a time when the profession was dominated by Popper's (1965) view that there is no logic of discovery. Buchanan stated the new research

program: "The traditional problem of finding an effective method for formulating true hypotheses that best explain phenomena has been transformed into finding heuristic methods that generate plausible explanations. The problem of giving rules for producing true scientific statements has been replaced by the problem of finding efficient heuristic rules for culling the reasonable candidates for an explanation from an appropriate set of possible candidates" [and finding methods for constructing the candidates] (Buchanan 1985, 110-111). Discovery as heuristic search in a search space enabled AI methods to be applied to discovery tasks.

The first expert system, DENDRAL, was a scientific discovery system. It formed hypotheses about chemical compounds, given mass-spectrographic data (Lindsay, Buchanan, Feigenbaum, & Lederberg, 1980; 1993). This was followed by Meta-DENDRAL, which discovered new rules in mass spectrographic analysis, so as to by-pass the problem of getting rules from experts (Buchanan & Feigenbaum, 1978). Although its original algorithm was a computational realization of Lederberg's systematic scan strategy (Lederberg, 1965), DENDRAL was built to carry out a contemporary, difficult scientific task rather than as a model of human cognition.

A more historical-cognitive approach was the aim of the work on BACON, which rediscovered various scientific laws by finding patterns in numerical data (Langley, Simon, Bradshaw & Zytkow, 1987). Simon's early work on finding patterns in sequences (Simon & Kotovsky, 1963) was extended in BACON to heuristic search for patterns in numerical data. The most creative of BACON's abilities was the decomposition of relational data to conjecture intrinsic properties in one or more of the objects engaging in the relations. This step went beyond curve-fitting and was based on the metaphysical assumption that an entity's relational properties are caused by its intrinsic properties. In addition to the data-driven tasks modeled in BACON, the group also investigated theory-driven discovery in STAHL. One wonders to what extent these programs model actual cognitive processes of historical scientists, as opposed to finding strategies which are sufficient to reproduce the historical results. As with most simulations, they provide "how possibly" accounts. Using studies of notebook evidence, the KEKADA system (Kulkari & Simon, 1988) modeled reasoning patterns in some discoveries of the biochemist Hans Krebs and focused on responses to surprising experimental results, helping to dispel the mystery of serendipity in discovery.

A seminal conference on computational methods for scientific discovery, whose proceedings were published in 1990 (Shrager & Langley, Eds.) is a useful source for the state to the field at that time.

Recent Work

Some of the pioneers in scientific discovery, e.g., Buchanan, Simon, and Zytkow, push ahead with their research programs. Others who contributed to the 1990 volume are still working on discovery. The American Association for Artificial Intelligence sponsored a Spring Symposium on Systematic Methods of Scientific Discovery in March, 1995. A special issue of *Artificial Intelligence* on computational discovery is about to appear, although fewer papers were received than the editors wished (Simon, Valdes-Perez & Sleeman, forthcoming). Data-mining in scientific databases is an active area of research, as are other computational approaches applied to individual sciences, e.g., intelligent systems in molecular biology. It is becoming more difficult to locate computational discovery work because much of it is published in scientific journals--a good sign that the methods of producing results of interest to practicing scientists.

Buchanan (e.g., Lee et al., 1996) continues work on rule induction applied to various scientific databases. Simon is studying the difficult problems of constructing diagrammatic representations (Larkin & Simon, 1987; Qin & Simon, 1995) and of modeling relations between diagrammatic and verbal reasoning (Tabachneck-Schijf, Leonardo, & Simon, 1996). Zytkow continues to work on various aspects of discovery, including analyzing the components needed for an autonomous discovery agent (e.g., Zytkow, 1995/96) and knowledge discovery in databases (e.g., Zytkow & Zembowicz, 1996).

Much of the current work in computational discovery is occurring within applications to particular sciences. According to Peter Karp, the whole field of bioinformatics is doing computational scientific discovery but there is a gradient from computational discoveries that are not based on AI methods, to computational discoveries that are based on AI methods, to methods with a "cognitive flavor." Not much of the bioinformatics work falls into the last category. However, Karp (et al., 1996) applied reasoning by analogy to predict metabolic pathways in the bacterium, *H. influenzae*, based on the extensive knowledge base that he and Monica Riley, a bacterial geneticist, have developed for *E. coli*.

Larry Hunter, a frequent editor of publications in AI and molecular biology (e.g., Hunter 1993), recently informed me that there is a clear success is the application of AI technology to molecular biology: hidden Markov models (HMMs) for molecular sequence analysis. They are being applied to automatically build models of families of nucleotide and amino acid sequences. These models are useful as extremely sensitive classifiers of novel sequences, and also generate multiple sequence alignments of large numbers of sequences in a computationally efficient way. Tools based on this approach are now in wide use in the biological community. A review article is Eddy (1996). Also, AI-based qualitative reasoning technologies have produced several good applications in reasoning about metabolism. Perhaps somewhat surprising is that the work in intelligent systems in molecular biology, for the most part, does not employ discovery methods discussed at the Shrager and Langley (Eds. 1990) conference.

The extensive protein sequence database has provided a challenge for those seeking to find computational methods to predict how the linear amino acids will fold into the secondary and tertiary structures in proteins. The Human Genome Project, which is rapidly producing millions of bases of sequence information about both human and model organism genomes, presents a challenge for computational approaches. Good programs are needed for discovering genes, both coding regions and regulatory regions, in these linear sequences. Current programs are not good at finding introns, intervening sequences between the coding regions of genes. Since the genetic system has some means of detecting introns, one can expect computational systems to be able to discover the signal(s). Knowledge discovery in scientific databases (e.g., Fayyad, Haussler & Stolorz, 1996) promises to be an important area in coming years.

Raul Valdes-Perez's (1994) work in chemistry shows the power of computational systems in doing a systematic search of a hypothesis space, given certain constraints. MECHEM is able to find reaction pathways that chemists have missed.

Buchanan's work on rule discovery in scientific databases and Valdes-Perez's work on systematically conjecturing chemical reaction pathways illustrate the power of design AI systems that aim, not at realistically modeling human cognitive capacities, but using computational methods to circumvent human limitations. Humans are not good at searching massive databases and manipulating sets of rules with many features to make predictions. Cognitive science research has shown that humans have a tendency to focus too rapidly on one hypothesis before doing a systematic search of a hypothesis space. Discovery programs that are more systematic and more thorough than humans are an aid to scientists.

Computational Discovery: Pros and Cons

My own work on reasoning in scientific change focuses on an cyclic process: discovery, assessment, revision. Given a good revision procedure, one's discovery methods can be weaker. Strategies for these processes include: strategies for producing new ideas, e.g., analogies, abstraction instantiation, interfield relations; strategies for theory assessment, e.g., prediction-testing, relations to theories in other fields; and strategies for anomaly resolution (Darden 1991, Ch. 15). After extensive historical study of the development of Mendelian genetics, I proposed hypothetical strategies of the three types. The historical evidence was inadequate to show that they are descriptive cognitive strategies actually used by geneticists. Instead, they are hypothetical strategies that *could* have been used in the historical development of the theory of the gene to produce the changes that *did* occur (Darden, 1991). One needs to show that these strategies are *effective* problem-solving strategies, instances of useful "compiled hindsight" (Darden, 1987), applicable to additional cases, worthy of being used by contemporary scientists or to build AI discovery systems.

I visited in Joshua Lederberg's Laboratory for Molecular Genetics and Informatics and participated in episodes of anomaly resolution that exemplified some of the revision strategies I had proposed (Darden & Cook 1994). One

difficulty with the live-in-the lab approach is that little may happen while you are there; fortunately, I was able to observe some anomaly resolution strategies in use. Although I have attempted to implement some of the strategies in AI programs in order to demonstrate their efficacy (e.g., Darden & Rada, 1988; Kettler & Darden 1993; Darden, 1997), I have returned to historical-philosophical work, testing whether strategies from the Mendelian case apply to molecular biology (Darden, 1995).

Computational discovery work has advantages and disadvantages. Finding an adequate knowledge representation for a scientific case is difficult. Early work attempted to represent the relations between genes and chromosomes in part-whole hierarchies and to implement reasoning via inheritance and upward propagation of properties (Darden & Rada, 1988). A much more fruitful method for knowledge representation in genetics was the functional representation (Josephsons, Eds., 1994) for genetic processes (Darden 1997). Furthermore, when one is designing a computational system to rediscover a historical hypothesis, one must navigate between designing a system that trivially reproduces exactly what one is seeking versus designing a system that is unable to accomplish the task at all. Analogy systems often suffer these problems: either the analog is represented in such a way that the system easily finds it or there are so many analogs that the task becomes impossible (for attempts to navigate between these problems, see Kettler & Darden, 1993; Holyoak & Thagard, 1995).

An advantage of computational methods is the precision and completeness that is required to build a working system. The philosopher-historian may neglect aspects that the programmer must specify in detail if the system is to run. A computational approach forces one to reexamine aspects that may be otherwise neglected. However, this advantage is purchased at the price of much time and effort to implement even small parts of a historical case. Various aspects of human discovery, such as the use of pictorial models (e.g., the beads on a string model for genes on chromosomes), provide substantial difficulties when designing an implementation. On the plus side, once one has invested the effort in building a running system, then there is the fun of running experiments, doing "what-if" analyses, testing alternative strategies.

The approach in our TRANSGENE system (Darden, Moberg, Thadani & Josephson, 1992; Darden, 1997) was also used by Karp (1990) in his GENSIM and HYPGEN systems and points to a fruitful way to design a computational discovery system. A qualitative simulator of biological (or other) processes is built and used to make predictions. Data is supplied to test the predictions and another component of the system compares the prediction with data, detects anomalies, and uses diagnosis/redesign strategies to localize the fault in the simulator and redesign a module to remove the anomaly. Perhaps this architecture may be of use in building future AI systems or perhaps more traditional simulation models might be coupled with a revision system to do diagnosis/redesign for anomaly resolution and model improvement.

It will be exciting to see what computational scientific discovery produces in the coming years.

Acknowledgments

The TRANSGENE work was supported by the General Research Board of the University of Maryland and the National Science Foundation Grant No. RII-9003142. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the National Science Foundation. The TRANSGENE system was designed in collaboration with John and Susan Josephson and Dale Moberg; TR.3 was implemented by Sunil Thadani. This paper was written while I enjoyed the hospitality of the Center for Philosophy of Science at the University of Pittsburgh. Very helpful were discussions with, and reprints received from, Bruce Buchanan and Herb Simon. Rapid email responses from Larry Hunter, Peter Karp, and Pat Langley were appreciated. Sets of reprints from Kevin Dunbar, Nancy Nersessian, Tom Nickles, and Jan Zytkow aided me in learning about their recent work. I enjoyed the demo of MECHEM by Raul Valdes-Perez and I profited from his web page: www.cs.cmu.edu/~sci-disc

References

- Alberdi, E. & Sleeman, D. (1997, forthcoming). "ReTAX: A Step in The Automation of Taxonomic Revision," *Artificial Intelligence* 91(2).
- Baker, L. M. & Dunbar, K. (1996). "Constraints on the Experimental Design Process in Real-World Science," in G.W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pp. 154-159. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bechtel, W. & Richardson, R.C. (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, N. J.: Princeton University.
- Buchanan, B.G. (1982). "Mechanizing the Search for Explanatory Hypotheses," in P. Asquith & T. Nickles (Eds.), *PSA 1982 V. 2*, pp. 129-146. Philosophy of Science Association, East Lansing, Michigan.
- Buchanan, B.G. (1985). "Steps Toward Mechanizing Discovery," in K. Schaffner (Ed.), *Logic of Discovery and Diagnosis in Medicine*. Berkeley: University of California Press.
- Buchanan, B.G. (1994). "The Role of Experimentation in Artificial Intelligence," *Phil. Trans. R. Soc. Lond. A* 349:153-166.
- Buchanan, B.G. & Feigenbaum, E. A. (1978). "DENDRAL and Meta-DENDRAL: Their Applications Dimension," *Artificial Intelligence* 11:5-24.
- Cohen, D. M., Kulikowski, C., & Berman, H. (1995). "DEXTER: A System that Experiments with Choices of Training Data Using Expert Knowledge in the Domain of DNA Hydration," *Machine Learning* 21:81-101.
- Corruble, V. & Ganascia, J.-G (1997, forthcoming). "Induction and The Discovery of the Causes of Scurvy: A Computational Reconstruction," *Artificial Intelligence* 91(2).
- Dama, M. & Dunbar, K. (1996). "Distributed Reasoning: An Analysis of Where Social and Cognitive Worlds Fuse,"

- in G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pp. 166-170. Mahwah, NJ: Lawrence Erlbaum Associates.
- Darden, L. (1976). "Reasoning in Scientific Change: Charles Darwin, Hugo de Vries, and the Discovery of Segregation," *Studies in History and Philosophy of Science* 7:127-169.
- Darden, L. (1980). "Theory Construction in Genetics," in T. Nickles (Ed.), *Scientific Discovery: Case Studies*. Dordrecht: Reidel.
- Darden, L. (1982). "Artificial Intelligence and Philosophy of Science: Reasoning by Analogy in Theory Construction," in T. Nickles & P. Asquith (Eds.), *PSA 1982*, V. 2, pp. 147-165. East Lansing, Michigan: Philosophy of Science Association.
- Darden, L. (1985). "Hugo de Vries's Lecture Plates and the Discovery of Segregation," *Annals of Science* 42:233-242.
- Darden, L. (1987). "Viewing the History of Science as Compiled Hindsight," *AI Magazine* 8(2):33-41.
- Darden, L. (1990). "Diagnosing and Fixing Faults in Theories," in J. Shrager & P. Langley (Eds.), *Computational Models of Scientific Discovery and Theory Formation*. San Mateo, California: Morgan Kaufmann.
- Darden, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.
- Darden, L. (1992). "Strategies for Anomaly Resolution," in R. Giere (Ed.), *Cognitive Models of Science*. Minneapolis: University of Minnesota Press.
- Darden, L. (1995). "Exemplars, Abstractions, and Anomalies: Representations and Theory Change in Mendelian and Molecular Genetics," in J. G. Lennox & G. Wolters (Eds.), *Concepts, Theories, and Rationality in the Biological Sciences*. Pittsburgh, PA: University of Pittsburgh Press.
- Darden, L. (1997). "Anomaly-Driven Theory Redesign: Computational Philosophy of Science Experiments," in T. W. Bynum & J. Moor (Eds.), *Digital Phoenix: How Computers are Changing Philosophy*. Oxford: Blackwell.
- Darden, L. (Ed.) (1996). *PSA 1996*, Proceedings of the 1996 Biennial Meeting of the Philosophy of Science Association, Part I: Contributed Papers, *Philosophy of Science*, Supplement to Volume 63, September 1996.
- Darden, L. (Ed.) (1997). *PSA 1996*, Proceedings of the 1996 Biennial Meeting of the Philosophy of Science Association, Part II: Symposium Papers, *Philosophy of Science*, Supplement to Volume 64, forthcoming September 1997.
- Darden, L. & Rada, R. (1988). "Hypothesis Formation Using Part-Whole Interrelations," in David Helman (Ed.), *Analogical Reasoning*. Dordrecht: Reidel.
- Darden, L. & Rada, R. (1988). "Hypothesis Formation Via Interrelations," in Armand Prieditis (Ed.), *Analogica*. Los Altos, California: Morgan Kaufmann.
- Darden, L., Moberg, D., Nagarajan, S. & Josephson, J. (1991). *Anomaly Driven Redesign of a Scientific Theory: The TRANSGENE.2 Experiments* (Tech. Rep. 91-LD-TRANSGENE). Columbus, OH: Laboratory for Artificial Intelligence Research, Ohio State University.
- Darden, L., Moberg, D., Thadani, S. & Josephson, J. (1992). *A Computational Approach to Scientific Theory Revision: The TRANSGENE Experiments* (Tech. Rep. 92-LD-TRANSGENE). Columbus, OH: Laboratory for Artificial Intelligence Research, Ohio State University.
- Darden, L. & Cook, M. (1994). "Reasoning Strategies in Molecular Biology: Abstractions, Scans and Anomalies," in D. Hull, M. Forbes, & R. M. Burian (Eds.), *PSA 1994*, v. 2, pp. 179-191. East Lansing, Michigan: Philosophy of Science Association.
- Dunbar, K. (1993). "Concept Discovery in a Scientific Domain," *Cognitive Science* 17:397-434.
- Dunbar, K. (1995). "How Scientists Really Reason: Scientific Reasoning in Real-World Laboratories," in R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge, MA: MIT Press.
- Dunbar, K. (1996). "Beyond the Myth of The Lone Scientist: Distributed Reasoning in Science and Scientific Discovery," *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* 3:24-26.
- Dunbar, K. (forthcoming). "How Scientists Think: On-line Creativity and Conceptual Change in Science," in T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Conceptual Structures and Processes: Emergence, Discovery and Change*. Washington, D. C.: American Psychological Association Press.
- Eddy, S. R. (1996). "Hidden Markov Models," from <ftp://genome.wustl.edu/pub/eddy/hmm-review.ps.Z>
- Fayyad, U., Djorgovski, S. G., & Weir, N. (1996). "From Digitized Images to Online Catalogs," *AI Magazine* 17 (2): 51-66.
- Fayyad, U., Haussler, D. & Stolorz, Z. (1996). "KDD for Science Data Analysis: Issues and Examples," In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp. 50-56. Menlo Park, CA: AAAI Press.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). "From Data Mining to Knowledge Discovery in Databases," *AI Magazine* 17 (3): 37-54.
- Griffith, T. W., Nersessian, N. & Goel, A. (1996). *The Role of Generic Models in Conceptual Change* (Tech. Rep. CIT-GOGSCI-96/02). Atlanta, GA: College of Computing, Georgia Institute of Technology.
- Holyoak, K. J. and Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*. Cambridge, Mass.: MIT Press.
- Hunter, L. (Ed.) (1993). *Artificial Intelligence and Molecular Biology*. Cambridge, MA: MIT Press.
- de Jong, H. & Rip, A. (1997, forthcoming). "The Computer Revolution in Science: Steps towards the Realization of Computer-Supported Discovery Environments," *Artificial Intelligence* 91(2).
- Josephson, J.R. & Josephson, S.G. (Eds.) (1994). *Abductive Inference: Computation, Philosophy, Technology*. New York: Cambridge University Press.
- Karp, P. (1989). *Hypothesis Formation and Qualitative Reasoning in Molecular Biology*. Ph.D. Dissertation, Stanford University, Stanford, California.
- Karp, P. (1990). "Hypothesis Formation as Design," in J. Shrager and P. Langley (Eds.), *Computational Models of*

- Scientific Discovery and Theory Formation*. San Mateo, California: Morgan Kaufmann.
- Karp, P. (1993). "A Qualitative Biochemistry and its Application to the Regulation of the Tryptophan Operon," in L. Hunter (Ed.), *Artificial Intelligence and Molecular Biology*. Cambridge, MA: AAAI Press and MIT Press.
- Karp, P., Ouzouni, C. & Paley, S. M. (1996). "HinCyc: A Knowledge Base of the Complete Genome and Metabolic Pathways of *H. influenzae*," in D.J. States, P. Agarwal, T. Gaasterland, L. Hunter & R. Smith (Eds.), *ISMB-96: Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology*, pp. 116-124. Menlo Park, CA: AAAI Press.
- Kettler, B. & Darden, L. (1993). "Protein Sequencing Experiment Planning Using Analogy," in L. Hunter, D. Searls, & J. Shavlik (Eds.), *ISMB-93, Proceedings of the First International Conference on Intelligent Systems for Molecular Biology*, pp. 216-224. Menlo Park, CA: AAAI Press.
- Kleiner, S. A. (1993). *The Logic of Discovery: A Theory of the Rationality of Scientific Research*. Dordrecht: Kluwer.
- Kulkarni, D. and Simon, H.A. (1988). "The Processes of Scientific Discovery: The Strategy of Experimentation," *Cognitive Science* 12:139-175.
- Langley, P., Simon, H.A., Bradshaw, G.L. & Zytkow, J.M. (1987). *Scientific Discovery: Computational Explorations of the Creative Process*. Cambridge, Massachusetts: MIT Press.
- Larkin, J. H. & Simon, H. A. (1987). "Why a Diagram is (Sometimes) Worth Ten Thousand Words," *Cognitive Science* 11:65-99.
- Lederberg, J. (1965). "Signs of Life: Criterion-System of Exobiology," *Nature* 207:9-13.
- Lee, Y., Buchanan, B. G., Klopman, G., Dimayuga, M. & Rosenkranz, H. S. (1996). "The Potential of Organ Specific Toxicity for Predicting Rodent Carcinogenicity," *Mutation Research* 358:37-62.
- Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A. & Lederberg, J. (1980). *Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project*. New York: McGraw Hill.
- Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A. & Lederberg, J. (1993). "DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation," *Artificial Intelligence* 61:209-261.
- Nersessian, N. J. (1992). "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science," in R. N. Giere (Ed.), *Cognitive Models of Science*. Minneapolis: University of Minnesota Press.
- Nersessian, N. J. (1995). "Opening the Black Box: Cognitive Science and History of Science," *OSIRIS* 10:194-211.
- Nickles, T. (Ed.) (1980). *Scientific Discovery, Logic and Rationality*. Dordrecht: Reidel.
- Nickles, T. (1994). "Enlightenment Versus Romantic Models of Creativity in Science--and Beyond," *Creativity Research Journal* 7:277-314.
- Popper, Karl (1965). *The Logic of Scientific Discovery*. New York: Harper Torchbooks.
- Provost, F. J. & Buchanan, B. G. (1995). "Inductive Policy: The Pragmatics of Bias Selection," *Machine Learning* 20:35-61.
- Qin, Y. & Simon, H.A. (1995). "Imagery and Mental Models," in J. Glasgow, H. H. Narayanan, & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Computational and Cognitive Perspectives*. Menlo Park, CA: AAAI/MIT Press.
- Richman, H. B., Staszewski, J. J. & Simon, H. A. (1995). "Simulation of Expert Memory using EPAM IV," *The Psychological Review* 102:305-330.
- Schaffner, K. (1993). *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.
- Schunn, C. D. & Dunbar, K. (1996). "Priming, Analogy, and Awareness in Complex Reasoning," *Memory & Cognition* 24:271-284.
- Shrager, J. & Langley, P. (Eds.) (1990). *Computational Models of Scientific Discovery and Theory Formation*. San Mateo, California: Morgan Kaufmann.
- Simon, H.A. (1977). *Models of Discovery*. Dordrecht: Reidel.
- Simon, H.A. and Kotovsky, K. (1963). "Human Acquisition of Concepts for Sequential Patterns," *Psychological Review* 70:534-546.
- Simon, H.A., Valdes-Perez, R.E. & Sleeman, D.H. (1997, forthcoming). Editorial: "Scientific Discovery and Simplicity of Method," *Artificial Intelligence* 91(2).
- Spirtes, P., Glymour, C. & Scheines, R. (1993). *Causation, Prediction, and Search*. New York: Springer Verlag.
- Swanson, D.R. & Smalheiser, D.H. (1997, forthcoming). "An Interactive System for Finding Complementary Literatures: A Stimulus to Scientific Discovery," *Artificial Intelligence* 91(2).
- Tabachneck-Schijf, H. J. M., Leonardo, A. M., Simon, H.A. (1996). *CaMeRa: A Computational Model of Multiple Presentations* (Tech. Rep. CIP 530). Pittsburgh, PA: Carnegie Mellon University.
- Valdes-Perez, R.E. (1992). "Algorithm to Generate Reaction Pathways for Computer-Assisted Elucidation," *Journal of Computational Chemistry* 13:1079-1088.
- Valdes-Perez, R.E. (1992). "A Necessary Condition for Catalysis in Reaction Pathways," *Journal of Physical Chemistry* 96:2394-2396.
- Valdes-Perez, R.E. (1993). "Algorithm to Test the Structural Plausibility of a Proposed Elementary Reaction," *Journal of Computational Chemistry* 14:1454-1459.
- Valdes-Perez, R.E. (1994). "Algorithm to Infer the Structures of Molecular Formulas within a Reaction Pathway," *Journal of Computational Chemistry* 15:1266-1277.
- Valdes-Perez, R.E. (1994). "Human/computer Interactive Elucidation of Reaction Mechanisms: Application to Catalyzed Hydrogenolysis of Ethane," *Catalysis Letters* 28:79-87.
- Valdes-Perez, R.E. (1994). "Heuristics for Systematic Elucidation of Reaction Pathways," *Journal of Chemical Information and Computer Sciences* 34:976-983.

- Valdes-Perez, R.E. (1994). "Conjecturing Hidden Entities via Simplicity and Conservation Laws: Machine Discovery in Chemistry," *Artificial Intelligence* 65:247-280.
- Valdes-Perez, R.E. (1995). "Generic Tasks of Scientific Discovery," in R. Valdes-Perez (Ed.) *Working Notes: AAAI Spring Symposium Series: Symposium: Systematic Methods of Scientific Discovery*.
- Valdes-Perez, R.E. (1995). "Some Recent Human/Computer Discoveries in Science and What Accounts for Them," *AI Magazine* 16(3):37-44.
- Valdes-Perez, R.E. (1995). "Machine Discovery in Chemistry: New Results," *Artificial Intelligence* 74:191-201.
- Valdes-Perez, R.E. (1995). "Machine Discovery Praxis," *Foundations of Science* 1:219-224.
- Valdes-Perez, R.E. (1996). "A New Theorem in Particle Physics Enabled by Machine Discovery," *Artificial Intelligence*, 82:331-339.
- Valdes-Perez, R.E. (1996). "Computer Science Research on Scientific Discovery," *Knowledge Engineering Review* 11:57-66.
- Valdes-Perez, R.E. & Minden, J.S. (1995). "Drosophila Melanogaster Syncytial Nuclear Divisions are Patterned: Time-Lapse Images, Hypothesis, and Computational Evidence," *Journal of Theoretical Biology* 175:525-532.
- Valdes-Perez, R.E. & Zytkow, J.M. (1996). "Systematic Generation of Constituent Models of Particle Families," *Physical Review E* 54:2102-2110.
- Zembowicz, R. & Zytkow, J.M. (1996). "From Contingency Tables to Various Forms of Knowledge in Databases," in U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusany (Eds.), *Advances in Knowledge Discovery and Data Mining*, Menlo Park, CA: AAAI Press.
- Zytkow, J.M. (1993). "Introduction: Cognitive Autonomy in Machine Discovery," *Machine Learning: Special Issue on Machine Discovery* 12:7-16.
- Zytkow, J.M. (1995/96). "Creating a Discoverer: Autonomous Knowledge Seeking Agent," *Foundations of Science* 2:253-283.
- Zytkow, J.M. (1996). "Automated Discovery of Empirical Laws," *Fundamenta Informaticae* 27:299-318.
- Zytkow, J.M. (Ed.) (1996). *Machine Discovery*. Kluwer Academic Publishers: Dordrecht.
- Zytkow, J.M. & Lewenstam, A. (1990). "Analytical Chemistry; The Science of Many Models," *Fresenius Journal of Analytical Chemistry* 338:225-233.
- Zytkow J.M., Zhu, J. & Hussam, A. (1990). "Automated Discovery in a Chemistry Laboratory," in *Proceedings of the Eighth National Conference on Artificial Intelligence*, pp. 889-894. Menlo Park, CA: AAAI Press.
- Zytkow, J.M. & Zembowicz, R. (1993). "Database Exploration in Search of Regularities," *Journal of Intelligent Information Systems* 2:39-81.
- Zytkow, J.M. & Zembowicz, R. (1996). "Automated Pattern Mining with a Scale Dimension," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 889-894. Menlo Park, CA: AAAI Press.