Automated Discovery in a Chemistry Laboratory*

Jan M. Żytkow† **Jieming Zhu** **Abul Hussam**
Department of Computer Science Department of Chemistry
zytkow@gmuvax.gmu.edu jzhu@gmuvax.gmu.edu
George Mason University, Fairfax, VA 22030 U.S.A.

Abstract

We describe an application of the discovery system FAHRENHEIT in a chemistry laboratory. Our emphasis is on automation of the discovery process as oposed to human intervention and on computer control over real experiments and data collection as opposed to the use of simulation. FAHRENHEIT performs automatically many cycles of experimentation, data collection and theory formation. We report on electrochemistry experiments of several hour duration, in which FAHRENHEIT has developed empirical equations (quantitative regularities) equivalent to those developed by an analytical chemist working on the same problem. The theoretical capabilities of FAHRENHEIT have been expanded, allowing the system to find maxima in a dataset, evaluate error for all concepts, and determine reproducibility of results. After minor adjustments FAHRENHEIT has been able to discover regularities in maxima locations and heights, and to analyse repeatability of measurements by the same mechanism, adapted from BACON, by which all numerical regularities are detected.

1. Introduction and motivation

Our current research on computer discovery systems has been guided by three long term goals. First, we are pursuing real, not simulated applications in a science laboratory. Second, we want to minimize human intervention in the working of our system, eventually aiming at a fully automated discovery. The third goal is generality of the discovery mechanism. We want to make our system useful to any experimental scientist dealing with acquisition and analysis of numerical data. Although much remains to be done, our discovery system FAHRENHEIT has made progress on all of these goals.

Discovery systems: a brief summary In the last dozen of years many computer programs have been constructed that simulate various aspects of scientific discovery. The earlier systems concentrated on qualitative and quantitative regularities, discovery of intrinsic concepts, the scope of the laws, and explanation of the hidden structure (Buchanan and Mitchell, 1978; Langley, 1978, 1981; Bradshaw, Langley, and Simon, 1980; Langley et al. 1983, 1987; Zytkow and Simon, 1986; Zytkow and Koehn, 1986; Falkenhainer and Michalski, 1986).

The growth of discovery systems has accelerated in the last three years. Several abilities lacking in earlier discovery systems have been introduced, primarily the ability to consider empirical context of a law (IDS: Langley and Nordhausen, 1990; GALILEO: Zytkow 1990; Sleeman, Stacey, Edwards, and Gray, 1989), the ability to design experiments (KEKADA: Kulkarni and Simon, 1987; FAHRENHEIT: Zytkow, 1987; Langley and Żytkow 1989), the ability to represent objects, states and processes (Langley and Nordhausen, 1990; Żytkow 1990), and the ability to reason by analogy (Falkenhainer, 1987; Falkenhainer and Rajamoney, 1988). Sleeman et al. (1989) suggested an interesting search in the space of qualitative models of a chemical system. All these new abilities have deepened our understanding of selected aspects of discovery, but significant progress has been made also on the important issue of integration. Two systems, both descendents of BACON, reached a considerable integration: IDS and FAHRENHEIT, the latter augmented by the GALILEO system that generalizes laws by decomposing them into simpler expressions.

Data acquisition problems. Virtually all discovery systems are simplistic in handling data acquisition and experimental error.

Error handling. Discovery systems were typically limited to 'clean' input data. Although many systems include some error-related parameters, they disregard the breadth and notoriety of real scientist's problems with raw data. For instance, a single error parameter for the whole system in BACON and ABACUS produces a funny result when a system cannot find a regularity for a variable V, but then it finds that V^4 is constant. This may happen when the values of V are

*The work described in this paper was supported by Office of Naval Research under grant No. N00014-88-K-0226

†On leave from Wichita State University.

between 1 and 0 because for each power of V they differ less and less with respect to the unchanging value of error. Experimental science not only requires that the empirical error is determined, but also that it is reduced as much as possible. This involves analysis of repeatability and detection of abnormal experimental conditions that invalidate a particular collection of data, such as a contaminated electrode.

Data acquisition bottleneck. As the discovery systems become more complex, they require more input data. A computer system can be provided with data in several ways. In the simplest variant, a human operator responds to the system's request for data by calculating and typing in the appropriate values. Some of BACON's original experiments were conducted in this way. When the number of requested experiments becomes large, it is preferred to use a simulation. Simulation is easy for toy problems, but it is difficult to build a simulator that would model a situation typical in a science laboratory where a variety of unexpected and unwanted phenomena influence the experiments, causing error of measurement and inducing sophisticated measurement procedures. If we want our discovery systems to deal with these effects, it is easier to study them in interaction with the real world.

Empirical semantics. The role of operational semantics, which links terms in scientific theories with direct observations and manipulations (Bridgman, 1927; Carnap, 1936; Żytkow, 1982) is negligible in the existing discovery systems, while it is very important in real science. Especially the use of instruments has been disregarded by research on machine discovery.

While the simplistic approach to experimentation was useful at the early stage of research on discovery systems, at the present time we may directly confront the main goal of science, which is developing theories of the real world in an interaction with the world. If discovery systems are going to satisfy the needs of the experimental scientist, we must spend more time working on real experiments and on processing real data. Because even a simple discovery in modern science is based on the analysis of many thousand datapoints, discovery systems should be ready to handle massive data rather than few datapoints on which we tested our early systems.

Automation in chemistry. Automation of the thinking processes which we are building into discovery systems must be matched by automation of the actual experimentation and data collection in the science lab. Are scientists getting ready for that? Actually, the automation and computer support in a science lab has already reached a massive scale, including data acquisition and data analysis. Recently the robot technology has been tested in the chemistry laboratory on the tasks of sample preparation, and in a relatively short time a number of complex lab operations can be automated by combination of simple steps (Sharp, Whitfield, and Fox, 1988).

To all for whom automation of discovery looks like a science fiction, we would like to deliver a practical proof that it is already possible on a limited scale. This paper reports our first results. But if it is possible to use a discovery system as a scientist's associate, does it pay to do so? It depends on the generality of the system. The BACON family of discovery systems, including BACON, IDS, and FAHRENHEIT, comes off very favorably on this account. Not only does each experimental scientist speak in terms of independent and dependent variables, empirical equations, their stepwise generalization, and their range of application, measurement error, relevant and irrelevant variables, but these terms refer to a larger part of their laboratory work. A reliable discovery system can become a workhorse that carries out a considerable amount of work in any science lab.

2. Chemistry experiments

We selected electrochemistry as the application area for a number of reasons. First, the computer operated sensors and actuators are readily available in a typical chemistry lab. Second, it is relatively easy to automate various electrochemical operations on chemical samples. Third, chemists often work with relatively simple equations, which do not exceed the theorizing capabilities of the systems in the BACON family.

We report on an automated experiment in differential pulse voltammetry. Our experiments parallel an ongoing research conducted by a chemist working on construction of new instrumentation. This gives us an opportunity to compare and verify our results.

Problem 1: Detection of low concentration ions. Differential pulse voltametry (DPV) can be used to measure concentration of ions such as Cu^{2+}, Pb^{2+}, and Cd^{2+}, simultaneously and at a very low concentrations of few particles per million (ppm) and lower (Bard and Faulkner, 1974). The problem is important since those ions are toxic at ppm and sub-ppm levels. DPV uses excitation of a sample by potential pulses characterized by the pulse height, width, and the delay between pulses (Figure 1a), and measures the response as a function of current in potential (Figure 1b). Figure 3 shows a typical result of a single series of experiments in which a particular pulse height, width, and delay between pulses have been used and three types of ions have been detected: cadmium, lead and copper. The type of an ion is determined by peak's location whereas concentration is proportional to peak's height. The location and height of each peak as well as the errors with which they are determined vary with the pulse parameters. Typically the resolution (discrimination between types of ions) is better at low values of pulse height, while the sensitivity is higher at larger values, that is, the concentration is measured with smaller error. When the values of pulse height become too high,

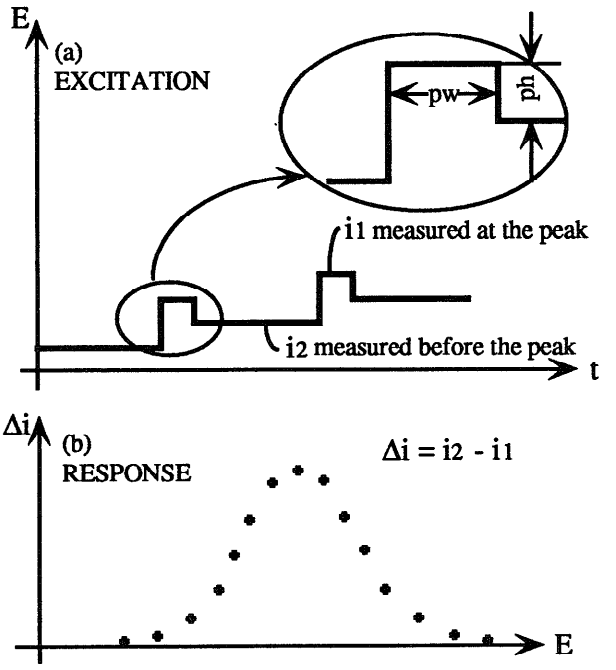

Fig. 1 Differential pulse voltammetry

(a) control signal -- voltage E as a function of time
(b) response -- difference in current Δi as a function of voltage E

however, the error in determining peak paramaters increases dramatically. To obtain the best resolution and sensitivity, not only the pulse height, but also the width, and the delay between pulses should be optimized. Instrumental to the optimization that would be efficient under different circumstances is a theory that describes how the peak parameters and their error depend on the pulse parameters.

The final objective for our FAHRENHEIT system on Problem 1 is to find such a theory in a fully automated way. In this paper we report on our initial results.

Problem 2: Peak repeatability. The initial phase of any scientific experiment is focused on experimental error and repeatability of results. The working electrode (Figure 2) poses repeatability problems because the measurements may deteriorate due to chemical changes at the electrode. To make the replacement of an electrode easy, chemists use mercury electrodes in which a hanging drop of mercury forms a working surface. When the performance of the electrode deteriorates, a drop of mercury can be forced out of the working electrode and replaced by a fresh, not contaminated drop. How many measurements can we perform with the use of the same drop? This is a typical repeatability problem, common for every sensor, and the procedure is also typical: perform a sequence of experiments trying to keep constant the values of all controllable parameters, and find the border between the area of repeatability, in which the results are constant

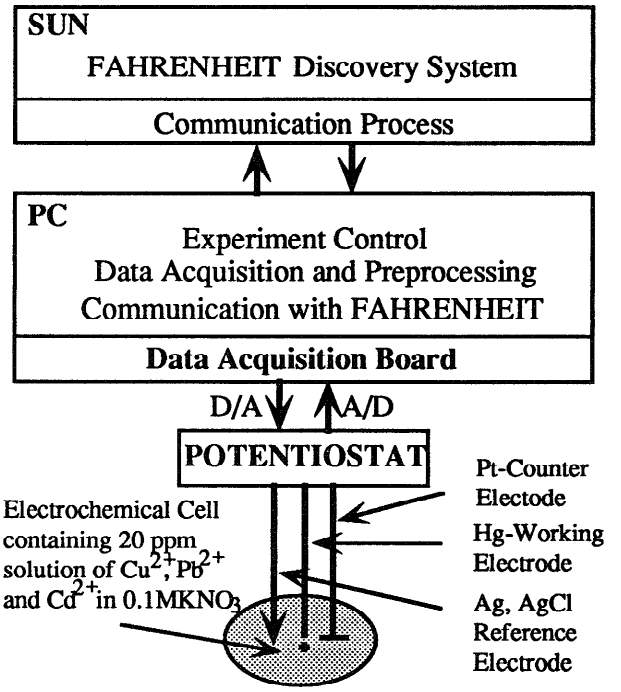

Fig. 2 Hardware and software configuration

within small error, and the area in which the results change or the error increases. By the analysis of the same sequence of experiments we must determine both the range of repeatability and the measurement error. Another source of error is the varying size of the drop. The amount of the current Δi is proportional to the surface of the drop. To determine the error we must experimentaly answer the question: "How accurately can we reproduce the drop size?"

Hardware and Software configuration Figure 2 illustrates hardware and software components and connections in our automated system, including the chemistry equipment. Our system runs a complete loop in which experiments are designed by FAHRENHEIT, performed under the control of PC in the electrochemical cell, the experimental results are sent to the Sun where FAHRENHEIT uses them to build a theory. Human interference is reduced to sample preparation and occasional assistance.

3. FAHRENHEIT's new features

FAHRENHEIT has been described by Koehn and Żytkow (1986), Żytkow (1987), and by Langley and Żytkow (1989). It uses BACON's mechanism for multidimensional control of experiments and for detection of multidimensional regularities, but it extends BACON in several directions. FAHRENHEIT searches for the scope of the laws it has discovered, it can find several regularities in data, change the order in which

it varies independent variables, and detect irrelevant variables.

Placing FAHRENHEIT "in charge" of laboratory experiments challenged us with new tasks. To be successful in the chemistry experiments, FAHRENHEIT must be able to (1) find "points of special interest", such as maxima, and incorporate them into a recursive mechanism for detection of multidimensional regularities, (2) determine reproducibility of results, (3) determine the error for each new concept, (4) establish empirical semantics, and (5) establish the communication between the brain (FAHRENHEIT) and the hands and eyes (PC), including a command language used by FAHRENHEIT. We will briefly describe the first four.

Handling "special points". Quantitative discovery systems were traditionally preoccupied with numerical regularities, whereas scientists are also interested in other aspects of numerical data, especially in "special points" such as maxima, minima, discontinuities, zeros, and so forth. Sometimes finding a special point is more important than detecting a regularity. FAHRENHEIT has been able to find many special points but it has used them for another purpose. FAHRENHEIT has been able to find boundaries of regularities, regularities on boundaries and boundaries of boundaries. It turned out that the same recursive mechanism for generalization to new variables applies to all types of special points. The system needed only a small change in code, after which it can mix and match various tasks of function analysis for different independent variables. For instance, FAHRENHEIT can now search for the maxima of regularities, for regularities on maxima, or for maxima of a regularity on minima. Problem 1 in section 2 requires detection of the location and height of each peak in a sequence of data for each pulse height, such as shown in Figure 3. Then, FAHRENHEIT is supposed to find separate regularities for the locations and heights of maxima for different ions. The electrode potential is the independent variable for the former, while pulse height is the independent variable for the latter.

Determining the reproducibility of results. In the repeatability study FAHRENHEIT performs a number of experiments without changing any values of controlled variables. Then it analyses the data, trying to separate an area of constancy from the remaining data in which there is either another regularity or no regularity can be found. FAHRENHEIT's capability for finding multiple regularities and their boundaries is essential on this task.

Handling experimental error. For each new variable generated by FAHRENHEIT, such as the maximum location or the slope of a linear regularity, the system finds the error. Each error is computed from appropriate data. For instance, the error of maxima location and height can be inferred from the repeatability study. The error is necessary for the generalization of results to new dimensions. It also allows to determine the quality of final results.

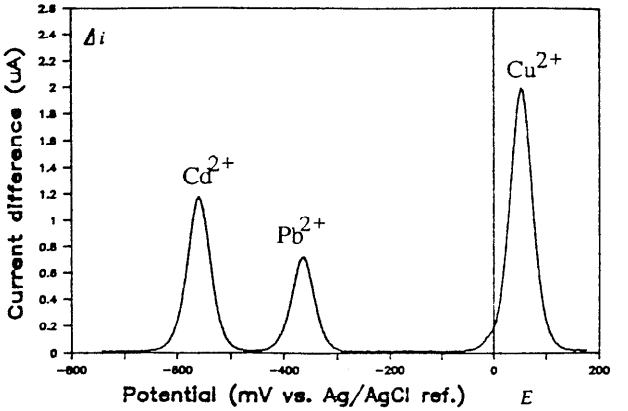

Fig. 3 DPV of metal ions (20ppm Cu^{2+}, Pb^{2+} and Cd^{2+})

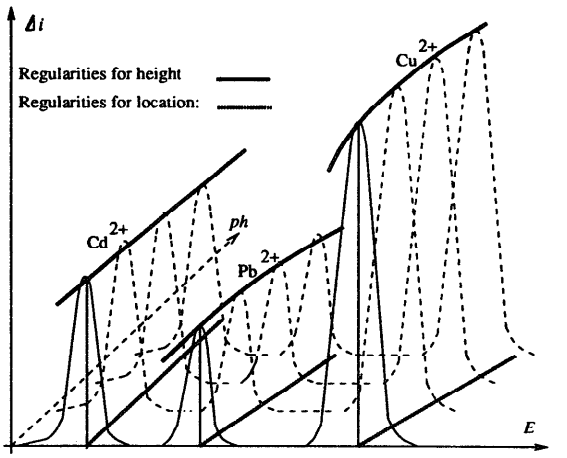

Fig. 4 Regularities for the maximum

4. Results of experiments

We started with the repeatability study, that is with problem 2 in section 2, using the pulse height of 15 mV. We repeated the experiment 20 times. Each time our system collected 360 datapoints, measuring the current as a function of potential of the working electrode. For a single pulse the results are plotted in Figure 3. In 120 minutes 7200 datapoints were collected and analysed, 60 maxima have been located and nine regularities, including six constancies, have been found. By analysing the ranges of constancy FAHRENHEIT detected that repeatability of maxima heights are different for different ions, and that each maximum is measured with a specific error as described in the "rows" errors and "repeatability" in Table 1. As a result, we use the number of 7 repetitions before we change the drop of mercury.

In the next experiment, reported in Figure 4 and in Table 1, we aimed at the detection of regularities that describe changes of the peaks as a function of the pulse

pulse h \ maximum	maximum1 Cd^{2+}		maximum2 Pb^{2+}		maximum3 Cu^{2+}	
	location	height	location	height	location	height
10	-554.90	0.51	-364.52	0.73	52.87	1.91
15	-554.90	0.57	-364.52	1.02	56.87	2.67
20	-551.24	0.62	-360.86	1.27	60.19	3.30
25	-554.90	0.67	-357.20	1.52	60.19	3.80
30	-551.24	0.71	-357.20	1.73	63.85	4.39
35	-543.90	0.72	-357.20	1.92	63.85	4.79
40	-543.92	0.77	-353.53	2.07	67.51	5.28
error/repeatibility	3.67/16	4%/5	3.67/20	4%/8	3.67/20	4%/5
regularity type	CONST	LIN	LIN	QUAD	LIN	QUAD
coefficient x*x	——	——	——	-0.0005	——	-0.001
coefficient x	——	0.01	0.36	0.072	0.50	0.17
coefficient const.	-553.87	0.4406	-368.45	0.063	47.6857	0.40

Table 1. Summary of DPV experiments. The upper part shows the parameters of the maxima. The first number in the row "error/repeatability" indicates the error, either absolute or relative(%), the second number indicates the range of repeatability. The lower part gives the type of the regularity and the numerical values of coefficients.

height. Figure 4 illustrates a sequence of experiments with a varying pulse height. FAHRENHEIT found the locations and heights of the maxima, reported in the upper portion of Table 1, then it found regularities both for peak locations and heights, depicted in Figure 4 and described in the lower portion of Table 1, where the type of regularity is given and the numerical values of the coefficients for each regularity. The experiment lasted 50 minutes, during which for seven pulse heights the total of 2320 datapoints were collected and analysed, 21 maxima have been located and six regularities have been found.

5. Evaluation

System's performance can be evaluated according to accuracy, range of application, and usefulness. The accuracy of our results is compatible with the accuracy achieved by human researchers. When we were able to compare the results, our system finds both the maxima and regularities which are equivalent within empirical error.

The range of applications of FAHRENHEIT goes far beyond electrochemistry, because FAHRENHEIT can deal with numerical data regardless of the application domain, provided that the laboratory hardware is rearranged and the appropriate operational procedures are defined.

Is our system useful? How fast it performs? The performance is still subject to many improvements, but the first results are encouraging. When we compare the theoretical work on the data made by FAHRENHEIT with an off-line analysis by the chemists, the results are very favorable. More than one day of work on data by a research assistant corresponds to 1 minute of data analysis by FAHRENHEIT. Altogether, several days work of research assistants corresponds to 50 minutes work of our system. The cost of the system installation is still enormous, measured in man-months of work, and it will take some time before we reach satisfactory portability.

6. Limitations and Future Directions

Limitations. Although FAHRENHEIT can automatically determine the repeatability conditions and it can find regularities for the peaks, the transition from the first to the second requires our intervention. We would like our system to understand the notion of repeatability, so it would investigate repeatability automatically whenever necessary, and it would be able to automatically use the results. The results of repeatability study influence the experimentation because they impose a limit after which the experiments can continue only after the electrode has been changed. The corresponding change in operational procedure, however, has not been automated and must be done manually.

Immediate future steps. We are planning to work on these limitations, and then to expand the theory of DPV measurements to additional independent variables, including pulse width, delay between pulses, and concentration of ions in the sample.

7. Conclusions

Our preliminary results demonstrate how a quantitative discovery system can be used in a chemistry laboratory on the experimental problems of interest to a contemporary chemist. Both an experimental scientist and a machine learning researcher can benefit from the application of a discovery system in a science laboratory. The scientist may save enormously on time and effort spent on data analysis, while the ML researcher may gain a valuable exposure to problems of real world experimentation and real data analysis.

FAHRENHEIT is superior over the existing statistical packages from the perspective of theory generation in that it can infer multidimensional regularities and it integrates many discovery steps. Various software packages are available that can help scientists in data processing, but they only automate single small steps, like smoothing and regression. The result of

each step must be inspected by a scientist before he decides about the next step and prepares the input for that step. In contrast, FAHRENHEIT automates large portions of the inference process.

Another advantage is a flexible data collection. Many chemistry labs use automated procedures for data collection which vary the independent variables in a predetermined pattern. These procedures do not provide automated feedback from the theoretical results, because theoretical analysis is done *after* empirical data have been collected. In contrast, FAHRENHEIT builds the theories *on-line* and it changes the experimentation pattern to concentrate new experiments on a particular area that best contribute to the growth of the emerging theory.

Researchers on discovery can also benefit because a real lab interface allows to overcome the bottleneck of data generation, provides realistic raw data, and yields new research problems such as on-line refinement of experimental procedures.

References

Bard, A.J., and Faulkner, L.R. 1974 *Electrochemical Methods: Fundamentals and Applications*, Wiley: New York.

Bradshaw, G., Langley, P., and Simon H.A. 1980. BACON 4: The Discovery of Intrinsic Properties. *Proceedings of the Third National Conference of the Canadian Society for Computational Studies of Intelligence*, 19-25.

Bridgman, P.W. 1927. *The Logic of Modern Physics.* Macmillan, New York.

Buchanan, B.G., and Mitchell, T.M. 1978. Model-directed learning of production rules, in: Waterman A., Hayes-Roth F. eds., *Pattern-Directed Inference Systems*, Academic Press, New York.

Carnap, R. 1936. Testability and Meaning, *Philosophy of Science, 3.*

Falkenhainer, B.C. 1987. Scientific Theory Formation Through Analogical Inference, *Proceedings of Fourth International Workshop on Machine Learning*, Los Altos, California: Morgan Kaufmann Publ., 218-229.

Falkenhainer, B.C., and Michalski, R.S. 1986. Integrating Quantitative and Qualitative Discovery: The ABACUS System, *Machine Learning, 1*, 367-401.

Falkenhainer, B.C., and Rajamoney, S. 1987. The Interdependencies of Theory Formation, Revision, and Experimentation, *Proceedings of the Fifth International Conference on Machine Learning.* Morgan Kaufmann Publ.: Los Altos, California, 353-366.

Koehn, B., and Zytkow, J.M. 1986. Experimenting and Theorizing in Theory Formation. In Ras Z. ed., *Proceedings of the International Symposium on Methodologies for Intelligent Systems* 296-307. ACM SIGART Press.

Kulkarni, D., and Simon, H.A. 1987. The Processes of Scientific Discovery: The Strategy of Experimentation, *Cognitive Science, 12*: 139-175.

Langley P.W. 1978. BACON 1: A General Discovery System, *Proceedings of the Second National Conference of the Canadian Society for Computational Studies*, 173-180.

Langley, P.W. 1981. Data-driven Discovery of Physical Laws. *Cognitive Science, 5*: 31-54.

Langley, P., Zytkow, J.M., Bradshaw, G., and Simon, H.A. 1983. Three facets of scientific discovery, *Proceedings of the Eighth International Joint Conference on Artificial Intelligence.*

Langley, P., and Nordhausen, B. 1986. A framework for empirical discovery. *Proceedings of the International Meeting on Advances in Learning.* Les Arc, France.

Langley, P.W., Simon, H.A., Bradshaw, G., and Zytkow J.M. 1987. *Scientific Discovery; An Account of the Creative Processes.* Boston, MA: MIT Press.

Langley, P.W., and Żytkow, J.M. 1989. Data-Driven Approaches to Empirical Discovery. *Artificial Intelligence, 40*: 283-312.

Nordhausen, B., and Langley, P. 1990. An Integrated Approach to Empirical Discovery. in: J.Shrager and P. Langley eds. *Computational Models of Discovery and Theory Formation*, Morgan Kaufmann Publishers, San Mateo, CA. in press.

Sharp, R.L., Whitfield, R.G., and Fox, L.E. 1988 Robotics in the Laboratory: a Generic Approach, *Analytical Chemistry, 60*: 1056-1062.

Sleeman, D.H., Stacey, M.K., Edwards, P., and Gray, N.A.B. 1989. An Architecture for Theory-Driven Scientific Discovery, *Proceedings of EWSL-89.*

Zytkow, J.M., and Simon, H.A. 1986, A Theory of Historical Discovery: The Construction of Componential Models, *Machine Learning, 1*: 107-36.

Żytkow, J.M. 1982. An interpretation of a concept in science by a set of operational procedures. *Polish Essays in the Philosophy of the Natural Sciences* 169-185. Dordrecht: Reidel.

Zytkow, J.M. 1987. Combining many searches in the FAHRENHEIT discovery system. *Proceedings of Fourth International Workshop on Machine Learning* 281-287. Los Altos, CA: Morgan Kaufmann.

Zytkow, J.M. 1990. 'Deriving basic laws by analysis of processes and equations', in: J.Shrager and P. Langley eds. *Computational Models of Discovery and Theory Formation*, Morgan Kaufmann Publishers, San Mateo, CA. in press, 1990.