Schunn, C. D., & Klahr, D. (1995). *A 4-space model of scientific discovery*. In Proceedings of the 17th Annual Conference of the Cognitive Science Society.

A 4-Space Model of Scientific Discovery

Christian D. Schunn
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
schunn@cmu.edu

David Klahr
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
klahr@cmu.edu

Abstract

An extension of Klahr and Dunbar's (1988) Dual space model of scientific discovery is presented. We propose that, in addition to search in an experiment space and a hypothesis space, scientific discovery involves search in two additional spaces: the space of data representations and the space of experimental paradigms. That is, discoveries often involve developing new terms and adding new features to descriptions of the data, and the also often involve developing new kinds of experimental procedures. The 4-space model was motivated by the analysis of human performance in a discovery microworld. A brief description of the data is presented. In addition to the general 4-space framework, a description of the component processes involved in each of the four search spaces is also presented.

Overview

One fruitful characterization of scientific discovery is to view it in terms of search in two problem spaces: a space of hypotheses and a space of experiments (Klahr & Dunbar, 1988; Simon & Lea, 1974). This characterization can be used to classify discovery models into three groups. First, there are those that address the processes of hypothesis generation and evaluation (e.g., the BACON models (Langley, Simon, Bradshaw, & Zytkow, 1987), COPER (Kokar, 1986), and ECHO (Thagard, 1988)). Second, there are those that address the process of experiment generation and evaluation (e.g., DEED (Rajamoney, 1993), and DIDO (Scott & Markovitch, 1993)). Third, there are those that address both processes (e.g., KEKADA (Kulkarni & Simon, 1988), STERN (Cheng, 1990), and SDDS (Klahr & Dunbar, 1988)).

Based on our analysis of subject performance in a complex computer microworld, we have extended the 2-space framework to a 4-space framework. In the new framework, what was previous conceived as the hypothesis space has now been divided into a *data representation space* and a *hypothesis space*. In the data representation space, representations or abstractions of the data are chosen from the set of possible features. In the hypothesis space, hypotheses about causal relations in the data are drawn using the set of features in the current representation. Similarly, the old experiment space is now divided into an *experimental paradigm space* and an *experiment space*. In the experimental paradigm space, a class of experiments (i.e., a paradigm) is chosen which identifies the factors to vary, and the components which are held constant. In the experiment space, the parameters settings within the selected paradigm are chosen.

We made these changes as we began to scrutinize the human performance data from several discovery microworlds in preparation for the computational implementation of the 2-space model. It became clear that, during the course of their investigations of the domain, subjects often acquired new data representations, and developed new kinds of experiments. Furthermore, representation and paradigm selection appear to require different mechanisms from those necessary for hypothesis and experiment selection.

Our goal is to produce a model of processing in all 4 problem spaces. This model will consist of separate components corresponding to processing in each of the four problem spaces. However, as indicated in Figure 1, processing within each space is dependent upon the current state of the search in the other spaces. For example, experiment space search depends upon the available experimental paradigms as well as the current hypothesis. The arrows between the four spaces are those implied by the processes that we have found to exist—others connections may also exist. Given these strong interdependencies, there is great advantage to implementing each of the components in a unified model.

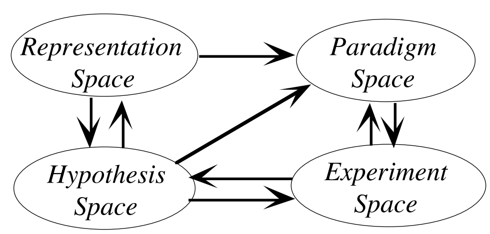

Figure 1: Information flow between the 4 search spaces.

Before presenting the details of the model, we will provide a brief description of the task and data that lead to the new model.

The Discovery Task

The task that contributed the data for our model design is a complex computer microworld called MilkTruck (Schunn & Klahr, 1992, 1993), in which subjects conducted experiments to discover the action of a complex mystery

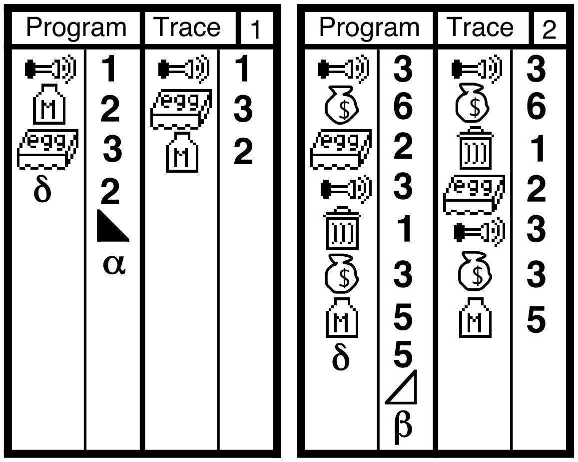

Figure 2: Two example programs and outcomes.

function. In the microworld, a "milk truck" executed a sequence of actions associated with a dairy delivery route. At any of 6 different locations along its route, it could beep its horn, deliver milk or eggs, or receive money or empties. A program consisted of a sequence of up to 14 action-location pairs. After the route had was entered, the subject pressed 'RUN' and the milk truck executed its route on the screen. The milk truck went to each location on the programmed route in the order that it was programmed, and animated icons demonstrated what transpired at each location.

In this task, subjects were given a great deal of external memory support. As subjects entered their programs, the steps were displayed on the screen in the program listing. Also, as the route was completed, a trace listing displayed in program format what transpired during the run (see figure 2). The subjects were also given access to all previous programs and traces.

The subject's task was to discover the function of a mystery command called δ (delta), which was a complex function with three arguments: a number (1 – 6), a triangle (white or black), and a Greek letter (α or β). When δ was not used, the trace listing was identical to the program listing. However, δ could change the order of delivery, and the resultant route execution and its associated trace would then be discrepant from the program listing. The effect of δ was to reorder the execution of part of the program according to the values of its three arguments (see Table 1).

Table 1: For the last N steps in the program, δ reorders the execution sequence of the program by...

	◿ (increasing)	◢ (decreasing)
α (item)	...item in increasing keypad order.	...item in decreasing keypad order.
β (house)	...house in increasing number order.	...house in decreasing number order.

The subjects were Carnegie Mellon University undergraduates. Subjects typically took part in a single, 1 hr session. Following an introduction to the basics of the MilkTruck domain, the syntax of δ was described, and the goal of discovering the effect of δ was presented to the subjects. In the discovery phase, subjects designed, conducted, and analyzed experiments with the goal of discovering the role of δ and its arguments.[1] The subjects worked at the discovery task until they had solved it, or they wished to give up.

The Data

Data was collected from over 100 subjects across various conditions. Both key-stroke and verbal protocols were collected. Here, we present a very brief description of the characteristics of subjects' behavior that led us to the creation of the 4-space model. In particular, we will focus on the evidence which suggested the addition of the experimental paradigm and data representation spaces. Data motivating the details of the experiment and hypothesis space processes can be found elsewhere (e.g., Schunn & Klahr, 1992, 1993, 1995).

The primary evidence for activity in the data representation space involved changes in subjects' descriptions of experimental outcomes. Early in the sessions, subjects typically described experimental outcomes in terms of series of movements of single steps. For example, with program 2 of figure 2, subjects early on in the discovery session would describe this outcome as follows: the third step moved to the fourth position, the fourth step moved to the fifth position, and the fifth step moved to the third position. Later in the sessions, subjects began to give descriptions for, and hypothesize about, the same kinds of experimental outcomes in terms of movements of segments of the program. For example, with program 2 of figure 2, subjects later on in the session would describe this outcome as follows: the last five steps were reorganized by increasing house number. While one might argue that these changes were merely redescriptions or reorganizations of the same features, subjects also added completely novel features to their descriptions (e.g., the number of times the milk truck changed directions during the route; the number of times the milk truck driver jumped up and down at the end of the route). This kind of evidence led us to hypothesize that subjects changed the way in which the basic data was used by adding and deleting features to their data representations.

The primary evidence for activity in the experimental paradigm space derived from subject statements about their plans for experiment selections and changes in these plans over the course of the problem-solving session. Initially, subjects had very few kinds of experiments from which to select. Their typical programs simply involved selecting a

[1] In this context, a program is an "experiment" and a statement about how the parameters work is a "hypothesis".

small number of houses and items without further constraints or forethought (e.g., program 1 of figure 2). Later in the session, subjects began to develop more complex kinds of experiments. For example, a subject might design a program of the following type: a long program with two more steps than the δ number argument, houses and items all different and not in order (e.g., program 2 of figure 2). Subjects also developed multi-program paradigms. For example, a subject might decide to conduct a sequence of five programs with the same base program, varying only the δ number parameter. Subjects learned to generate these complex, very deliberately chosen experiments quite rapidly, indicating that their were choosing experiments from a newly compiled database of experiment types.

The Model

The 4-space model consists of more than just the four search spaces; there are also the constituent search processes within each space (see Table 2). A brief description of the processes that we have found to occur within each of the spaces is presented below (although there are likely to be many more than this). These descriptions also serve to further illustrate the relationships between the four spaces.

Table 2: The component processes within each of the search spaces of the 4-space model.

Space	Process
Experimental Paradigm	hypothesis testing analogy error analysis rep/ hyp change
Experiment	theory orientation complexity management risk regulation
Data Representation	notice-invariants analogy brute-force search
Hypothesis	piecemeal induction representational mapping pop-out

The Experimental Paradigm Space

On occasion, making an important discovery involves finding a new method for gathering data—a new experimental paradigm.[2] It is unlikely that his new method for gathering data is some new domain-general induction method (e.g., Mill's inductive cannons); instead it is likely to be a method unique to that field of inquiry (e.g., changing the temperature in a particular order, instructing subjects in a particular way). These developments are typically not new instruments being developed (although they can be); rather they are typically new methods for using the same instruments. The issue at hand is how such new methods are created. Our model includes several domain-general heuristics for the creation of such new methods.

Paradigms are primarily created in the service of testing a hypothesis. The hypothesis embodies a set of assumptions about what features of the experiment are of interest. An experimental paradigm is created that emphasizes the features of interest. For example, to test the hypothesis that the number of steps in the route matter, the subject would create a paradigm in which the number of steps was an explicit feature of the paradigm. The corollary of this paradigm creation process is that paradigms are also created to de-emphasize features which are hypothesized not to matter (either by holding those features constant, or by removing them entirely from the experiment).

Paradigms can also be created through analogy to other paradigms. For example, subjects in the MilkTruck task developed the paradigm of holding delivery item constant from the paradigm of holding house number constant. These analogous, example paradigms can be ones acquired through observation, or ones generated by oneself in other situations. Paradigms may also be created by analyzing the cause of failed experiments. For example, if an experiment produces an ambiguous outcome, a new paradigm can be created to disambiguate the outcome. In the MilkTruck domain, many subjects ran one experiment in which two steps in the program were identical (e.g., program 2 of figure 2)—subjects noticed that this kind of experiment produces an ambiguous outcome and rarely ran that kind of experiment again. Furthermore, these new paradigms may be created through an error analysis of thought experiments rather than actual experiments.

The Experiment Space

Many costs and risks are associated with conducting experiments (e.g., mental effort, money spent, and potential loss of face for a failed experiments), and one practical goal of experimentation is to minimize these costs and risks. Experimentation also has theoretical goals related to the acquisition of information about the world. For example, it is desirable to design experiments relevant to the question at hand, with easily-interpreted and unambiguous results. How are these often conflicting goals achieved in particular experiments?

In our model, the theoretically oriented processes of experiment selection are achieved using two main heuristics: the *examination heuristic* and the *discrimination*

[2]The most popular use of the term "paradigm," typically associated with Kuhn (1970), refers to a much larger entity than we are considering. In fact, Kuhn used the word "paradigm" in two senses (which he acknowledges in the postscript of the second edition): the large scale paradigm of a whole field, and the smaller scale experimental paradigms that are used in particular experiments (e.g., the paired-associates paradigm, or Sperling's iconic memory paradigm). We will use the term to refer to experimental paradigms of the second, smaller, kind.

heuristic. The examination heuristic selects experiments which directly demonstrate the hypothesized effect. For example, a hypothesis about the behavior of acids in the presence of water leads to the selection of an experiment involving water. This tendency produces what has been called the +H test bias (Klayman & Ha, 1987) in rule discovery tasks: rules of the form "X's are a member of the concept" will lead to the selection of X's, rather than things that are not X's, to test the rule. The discrimination heuristic selects experiments which can discriminate among competing hypotheses under consideration. This heuristic is used only when multiple hypotheses are being considered. Therefore, there is no bias to select highly discriminating experiments (experiments which discriminate among many potential hypotheses) in the absence of multiple, specific, active hypotheses. However, risk regulation does take into account expected information content.

The practical goals of experiment selection are met through processes of *complexity management* and *risk regulation*. These experiments selection processes derived from the following phenomena in the MilkTruck domain. Firstly, using verbal protocol data, it was found that subjects choose shorter experiments when they were confused (unsuccessful at explaining experimental outcomes), and they choose longer experiments when they were highly confident (successful at explaining experimental outcomes). Secondly, using computer keystroke timing data, it was found that when recent experiments were easy to design (indexed by quick keystrokes), subsequent experiments were more likely to be longer. Conversely, when recent experiments were difficult to design (slow keystrokes), subsequent experiments were more likely to be shorter.

Complexity management involves regulating experiment design and interpretation complexity, where complexity is defined relative to the current state of understanding and experimental expertise. For example, longer experiments are more difficult to generate when few operators for generating long experiments exist, and the longer experiments are more difficult to interpret when the knowledge of relevant dimensions is small.

Risk regulation involves choosing experiments based on their perceived probability of producing an informative outcome. In many cases, this involves choosing between experiments which have a low probability of being successful, yet would be very informative if they are successful, and experiments which have a high probability of being successful even though they contain little potential information.[3] For example, conducting experiments which vary few features from the previous experiment are likely to behave exactly as predicted, whereas experiments in which many features have been varied have the potential of producing very novel results yet carry the risk of producing uninterpretable results.

[3] A successful experiment is one that can be meaningfully predicted or postdicted.

Complexity management and risk regulation are often in opposition. For example, more complex experiments are more likely to be informative, but are also much more difficult to generate and interpret. These two factors are combined to produce an expected utility, which determines the final experiment choice. The balance between complexity management and risk regulation varies with expertise. For example, with experience, longer programs become more easily generated and interpreted, and so, all subjects in the MilkTruck domain wrote longer programs towards the end of the problem-solving session.

The Data Representation Space

How does one choose or change a data representation? Finding the general solution to these questions is a difficult task because there is no known universal language for describing data representations, nor is there a known universal generator of representations. As a partial solution to these questions, we present three heuristics used for selecting representations from a previously existing repertoire.

In our model, data representation change occurs through the following mechanisms: *Notice Invariants*, *Analogy*, and *Brute-force search*. *Notice Invariants* works as follows. Experience with experimental outcomes within a domain leads to the noticing of certain regularities. New representations are chosen which emphasize these regularities. This behavior is exemplified in the MilkTruck task as subjects begin to notice that the first part of the program rarely changes. They then change their data representations to include changing and unchanging segments of the program.

Analogy produces representations by analogy to previously understood phenomena. For example, such analogies might include: computers are like programmable calculators, and atoms are like the solar system. The features used in the analogical source are applied to the analogical target. This process is similar to a categorization process. One kind of situation that triggers this process is the occurrence of salient, expectation-violating events, which force the recategorization of objects and events.

Brute-force search is a process of searching haphazardly through the set of possible representations of objects in the environment (i.e., by considering each object, and all the features and feature clusters of each object). This is the method by which subjects in the MilkTruck domain tended to add features to their data representations. The order of search may be constrained by the salience or availability of the possible representations. The process of brute-force search typically occurs when the individual believes that the current representation may not include the causally-predictive features.

The Hypothesis Space

In our model, the fundamental character of search in the hypothesis space is the piece by piece construction of hypotheses. This process, called *piecemeal induction*, was the method by which all subjects in the MilkTruck domain developed their hypotheses. In the first stage of piecemeal induction, a hypothesis is generated (either from memory or from data). Then, a scoping processes determines the generality or scope of the hypothesis. For example, a subject in the MilkTruck domain might hypothesize that the δ key reorders the last N steps only when a black triangle is selected (in contrast to concluding that the δ key reorders the last N steps no matter which triangle is selected). The dimensions used to form the scope are chosen from the current data representation. On each dimension, the most general scope value is preferred in the absence of counter-evidence.

With one or more particular hypotheses as input, abstraction processes generate more general hypotheses. For example, in the MilkTruck domain, subjects often abstract the hypothesis that the last N steps of the program are reordered from the particular hypotheses that there is no change with N=1, and the last two steps are changes with N=2. The number of particular cases that are sufficient to warrant a generalization is dependent upon expectations about the variability in the domain of study (which can be modified with experience).

There are many candidate mechanisms for the generation of hypotheses. In the domains that we have studied, two main processes have been found: *representational mapping*, and *pop-out*. *Representational mapping* is mapping of objects (e.g., the triangles in the MilkTruck task) onto actions or parts of actions (e.g., the order direction of the step rearrangement), where both the object and the actions (and action parts) are already in the current representation. Representational mapping is similar to a memory search. The representations in memory are searched for correspondences. The more complex the mapping (i.e., greater number of predicates), or the less salient the to-be-mapped feature, the lower the probability that the mapping will occur.

Representational mapping uses two heuristics: unique-function, and same-type. The unique-function heuristic favors mapping objects with no other known function onto actions or components of actions with no other known cause. The same-type heuristic favors mapping objects onto things of the same dimensionality. For example, binary object factors (e.g., black and white triangles) tend to be mapped onto inherently binary output factors (e.g., forward and reverse order). No actual experimental outcomes are necessary for representational mapping, since representational mapping can work with abstract schemata as well as particular objects. Therefore, this mechanism is typically used for generating initial hypotheses in the absence of evidence.

Pop-out occurs through automatic, categorization processes. When certain evidence presents itself, certain relationships are uniformly entertained. For example, exact similarity (whether coincidental or not) is automatically noticed. This automatic process is dependent upon representational factors. For example, if a feature is not encoded, no similarity involving that feature can be noticed.

Comparison to Previous Work

The details of our model are similar in many respects to other discovery models. The search space with the greatest degree of similarity is the hypothesis space. In particular, our piecemeal induction processes are very similar to the quantitative and qualitative rule induction processes of the BACON models (Langley, et al., 1987). FAHRENHEIT (Zytkow, 1987) is the intellectual precursor of our scoping processes.

The pop-out mechanism we use is a very generic computational principle. Many production systems models have domain-specific productions which immediately recognize and hypothesize about certain kinds of relations and correspondences. For example, KEKADA (Kulkarni & Simon, 1988) immediately recognizes mixed or additive effects given certain kinds of data. In another domain, STERN (Cheng, 1990) immediately recognizes power functions in quantitative data.

Turning to experiment space processes, there are no models of discovery that explicitly address the issue of complexity management. In contrast, several discovery systems have methods for ordering the experiment space search such that experiments likely to produce useful information are considered first. For example, AM (Lenat & Brown, 1984) focuses attention on concepts that produce novel results. In a similar fashion, DIDO (Scott & Markovitch, 1993) uses a curiosity heuristic which favors experiments testing the maximally uncertain part of the hypothesis. However, DIDO regulates whether experiment outcomes are considered further or ignored rather than regulating which experiments are conducted.

The examination principle is implicit in many models (e.g., LIVE (Shen, 1993), AM (Lenat & Brown, 1984), EURISKO, DIDO (Scott & Markovitch, 1993), and DEED (Rajamoney, 1993)), but explicit only in IE (Shrager, 1985). The discrimination principle is also taken from Shrager's IE model. However, there are similar principles in several other systems, including DEED (Rajamoney, 1993), and ABD-Soar (Johnson, Krems, & Amra, 1994).

Very few discovery systems create new experimental paradigms, and fewer still have considered this search space explicitly. STERN (Cheng, 1990) is one of the few such models. It has only one very simple paradigm creation mechanism. The most important difference between the paradigm construction in STERN and our model is that STERN creates new paradigms in order to try something new, whereas our model creates new paradigms because some feature of the new paradigm is desired.

Data representation change also has rarely been modeled. However, Kaplan's SWITCH (1989) presents a few heuristics for representation change, and they are different from the three heuristics explicitly postulated here (e.g., change grain size on failure, and pursue hot ideas). Furthermore, there are several programs capable of proposing new intrinsic properties, which might be construed as one form of data representation change. For example, BACON.4 (Langley et al., 1987) discovers the intrinsic property gravitational mass from the properties of force and distance by searching for constant relations among factors. There are also several kinds of conceptual hierarchy discovery programs that discovery new categories (i.e., new representations) by searching for feature invariance (e.g., Fisher's COBWEB (1987)). However, there has been little previous treatment of the process by which features and objects are deleted from the data representation nor of the process of adding completely novel features (rather than creating new features by combining existing features).

Conclusion

We have presented a general framework for understanding scientific discovery: the 4-space model of experimental paradigm, experiment, data representation, and hypothesis. This framework is a significant extension to the experiment and hypothesis space focus of the great majority of previous models of discovery, and we expect it to be applicable to many discovery domains.

We also have outlined the way in which searches in these four spaces interact. These interdependencies make it advantageous to consider all four spaces. In particular, previous models of discovery may have been trying to solve the difficult problems of data representation and experimental paradigm search in the process of dealing with hypothesis and experiment space issues, and may have been confounding separate issues in the process. By considering these issues as conceptually distinct factors, and by studying their interrelations, we may gain further insight into the modeling of scientific discovery processes.

The discovery task may be more computationally tractable by considering the experimental paradigm and data representation spaces explicitly. For example, rather than trying to consider all possible experimental paradigms while designing an experiment, it is easier to simply select from a small set of currently available experimental paradigms, and make the small number of decisions available in the selected paradigm. This set of experimental paradigms may be modified with experience in the discovery domain. Similarly, rather than trying to develop hypotheses using a very complete data representation containing all possible objects and features, simply select from the small number of objects and features in the current data representation. This data representation is also modified with experience in the domain. Thus, in both cases, very large search spaces are converted into several, much smaller search spaces.

As of yet, we have not discussed the control processes that coordinate search between the four spaces. Part of this coordination is driven by the sequential structure of the task: first experiments are created and run, then they are analyzed. The remainder of this coordination is driven by the logical relationships between the four spaces: experimental paradigms must be selected/created before experimental details are selected; and data representations must be selected before hypotheses are evaluated and modified. However, there are some exceptions to this simple scheme: occasionally experimental paradigms are evaluated for their effectiveness immediately after an experiment is conducted; and new data representations are occasionally created in response to a failure to develop a new hypothesis using the existing representation.

While our model has some features in common with other discovery models (although there are many novel features), the details of our model derive from detailed, on-line human performance data. This is in contrast to the majority of the existing discovery models that are motivated primarily by Artificial Intelligence goals or by historical analyses. It is interesting to note the similarities in underlying processes between our model and these other models despite the different modeling goals.

The goal of our future computational work will be to pursue the complete implementation of our model, and assess the tractability of our theoretical model, as well as its generalizability to other domains. Furthermore, we wish to match our model more precisely to the empirical data obtained from our studies with the MilkTruck task.

Acknowledgments

This work was supported by graduate scholarships from la Formation de Chercheurs et l'Aide à la Recerche and the Natural Sciences and Engineering Research Council to the first author, and by grants from NICHHD and the A.W. Mellon foundation to the second author.

References

Cheng, P. C.-H. (1990). *Modeling scientific discovery*. Unpublished doctoral dissertation, The Open University, Milton Keynes.

Kaplan, C. A. (1989). *SWITCH: A simulation of representational change in the Mutilated Checkerboard problem* (Tech. Rep. No. 477). Pittsburgh: Carnegie Mellon University, Department of Psychology.

Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, **12**, 1-48.

Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**(2), 211-228.

Kokar, M. M. (1986). Determining arguments of invariant functional descriptions. *Machine Learning*, **1**(4), 403-422.

Kuhn, T. S. (1970). *The structure of scientific revolutions*, 2nd edition. Chicago: University of Chicago Press.

Kulkarni, D. & Simon, H.A. (1988). The process of Scientific Discovery: The strategy of Experimentation. *Cognitive Science*, 12, 139-176.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational explorations in the creative process*. Cambridge, MA: MIT Press.

Lenat, D. B., & Brown, J. S. (1984). Why AM and EURISKO appear to work. *Artificial Intelligence*, **23**, 269-94.

Rajamoney, S. A. (1993). The design of discrimination experiments. *Machine Learning*, **12**, 185-203.

Schunn, C. D., & Klahr, D. (1992). Complexity Management in a Discovery Task. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*.

Schunn, C. D., & Klahr, D. (1993). Self- Vs. Other-Generated Hypotheses in Scientific Discovery. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*.

Schunn, C. D., & Klahr, D. (1995). Complexity management and risk regulation in discovery learning. In preparation.

Scott, P. D., & Markovitch, S. (1993). Experience selection and problem choice in an exploratory learning system. *Machine Learning*, **12**, 49-67.

Shen, W-M. (1993). Discovery as autonomous learning from the environment. *Machine Learning*, **12**, 143-165.

Shrager, J. C. (1985). *Instructionless learning: Discovery of the mental model of a complex device*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition*. Hillsdale, NJ: Erlbaum.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.

Zytkow, J. M. (1987). Combining many searches in the FAHRENHEIT discovery system. In Proceedings of the Fourth International Workshop on Machine Learning. Irvine CA, 281-87.