Machine Learning, 14, 313–320 (1994)

Extended Abstract

Children, Adults, and Machines as Discovery Systems

DAVID KLAHR KLAHR@ANDREW.CMU.EDU
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

Editor: D. Sleeman

Date Submitted: 17 November 1992

1. Introduction

This is a summary of recent research on the discovery process in which a common framework for both psychological studies and machine learning (ML) approaches to scientific discovery is described. Early interest in the psychology of science can be traced to Bruner, Goodnow, & Austin (1956), Wason (1960), and Simon (1966, 1973), among others. The state of the art as of a dozen years ago is summarized in Tweney, Doherty, and Mynatt (1981). The more recent resurgence of interest in the "cognitive science of science," particulary in ML can be attributed to Simon and his colleagues (Langley, Simon, Bradshaw & Zytkow, 1987; Kulkarni & Simon, 1988; Qin & Simon, 1990, Cheng & Simon 1992; Valdes-Perez, Simon & Murphy, 1992). However, beyond the projects that focus on the construction of computational models of the discovery process (c.f., Shrager & Langley, 1990), there is a substantial amount of psychological research on various aspects of scientific discovery (e.g., Chin & Brewer, 1992; Brewer & Samarapungavan, 1991; Dunbar, 1989; Dunbar & Schunn, 1990; Gholson, Standish, Neimeyer & Houts, 1989; Giere, 1988; Gorman, 1992; Holland, Holyoak, Nisbett & Thagard, 1986; Klayman & Ha, 1987; Kuhn, 1989; Schauble, 1990; Schauble, Klopfer & Raghavan, 1991; Sodian, Zaitchik & Carey, 1991; Vosniadou & Brewer, 1992).

2. Scientific discovery as dual search

Scientific discovery is a type of problem solving in which search takes place in two quite distinct spaces: an hypotheses space and an experiment space (Klahr & Dunbar, 1988). Each space represents accumulated knowledge, a set of operators, and a set of general heuristics for constraining search. Additionally, the process requires operators for mapping between the two spaces. The SDDS ("Scientific Discovery as Dual Search") framework summarizes the way in which these component processes are organized. Its three major components include the following:

1. *Searching the hypothesis space (H-space).* SDDS characterizes the process of generating new hypotheses as a type of problem solving search, in which the initial state consists of some knowledge about a domain and the goal state is a hypothesis that can account for most of that knowledge in a more concise or universal form.

2. *Searching the experiment space (E-space).* This involves search in a space that is only partially defined at the outset. Constraints on the search must be added during the problem solving process. One of the most important constraints is to produce experiments that will yield interpretable outcomes. One novel feature of SDDS is that it acknowledges the important and common role of collecting data, of running so called "experiments," in the absence of a clear theory. This is not the conventionally assigned role for experimentation, but it is an essential part of the discovery process.

3. *Evaluating evidence.* This involves a comparison of the predictions derived from a hypothesis with the results obtained from the experiment. Relevant features must first be extracted, potential noise must be suppressed or corrected, and the resulting internal representation must be compared with earlier predictions.

3. Empirical investigations of adults and children as discovery systems

In our empirical studies, we wanted to observe the discovery behavior of adults and children in situations that required coordinated search in both the experiment space and the hypothesis space, as well as the evaluation of evidence produced by subject-generated experiments.

We used a programmable device with a variety of function keys, and we trained subjects on all of the basic functions until they could reach a common criterion level of performance. In our original work the device was a computer controlled toy robot tank - called "BigTrak"; in our later studies it was a computer-based micro-world with the same functionality.

Once all subjects were equally familiar with this novel "scientific domain," we asked them to discover how a new function key ("RPT") worked by including it in the programs that they wrote for the device. Subjects were instructed to formulate hypotheses about the new function and run experiments to test those hypotheses. This required decisions about hypotheses and decisions about experiments. Subjects were never told whether or not they had actually discovered how the new function worked. They had to decide themselves when to terminate search.

One of our first studies (Klahr & Dunbar, 1988) contrasted the performance of Carnegie Mellon undergraduates with children between the ages of 8 and 11 years. The results showed striking difference in success rates, but similar performance in "higher level" measures of effort, time, and so on. Nearly all the children failed to discover the correct rule for the new function, although most of the unsuccessful children were sure they had discovered the correct rule, and they terminated their experimentation quite satisfied with their discovery. In contrast, nearly all of the adults discovered the correct rule. But it was not a trivial task for them. In fact, with respect to average time, number of hypotheses and number of experiments, the adults were not very different from the children. Differences in discovery processes lay at a deeper level.

Analysis of subjects' search in the H-space and the E-space revealed that, among adults, there were two distinct types of subjects with fundamentally different strategies: "Theorists" proposed the correct hypothesis before they had sufficient evidence to induce it, while "Experimenters" proposed the correct rule only after they had sufficient evidence (from experiments generated in the absence of any hypothesis) to induce the rule.

In subsequent studies (Klahr, Fay, & Dunbar, 1993), we focused on developmental differences in the heuristics used to constrain search in the experiment space. We knew that subjects at all ages shared domain-specific knowledge that biased them in the same direction with respect to the plausibility of different hypotheses. We expected both age and scientific training to reveal differences in the domain-general heuristics used to constrain search in the experiment space. Our goal was to investigate the extent to which prior knowledge—as manifested in hypothesis plausibility—influenced how people designed experiments and how they interpreted the results of those experiments.

We modified our procedure by giving subjects a suggestion about how the unknown function might work (the Given rule). However, it never really worked that way, but instead worked according to some other rule (the Actual rule). Both the Given and Actual could be either plausible or implausible. In some conditions, the Given hypothesis was only "somewhat" wrong, in that it was from the same frame as the way that the RPT key actually worked. In other conditions, the Given was "very" wrong, in that it came from a different frame than the Actual rule. We used four different groups of subjects: Carnegie Mellon (CM) undergraduates, Community College (CC) students, sixth graders (mean age 11 years), and third graders (mean age 9 years).

The correct (Actual) rule was discovered by 83% of the CMs, 65% of the CCs, 53% of the sixth graders, and 33% of the third graders. This group effect was attributable to conditions in which the actual rule was implausible: 56% of the adults but only 13% of the children were successful. In fact, none of the 3rd graders discovered implausible rules. (For a detailed analysis, see Klahr, Fay & Dunbar, 1993.)

In addition to overall success rates, we looked at a variety of measures that indicated how subjects responded to the experimental conditions. We analyzed how they responded to plausible and implausible Given hypotheses, and we looked at how they imposed constraints on search in the hypothesis space. The analysis revealed distinctive patterns of search, resulting from a set of domain-general heuristics that are differentially available to children and adults. Based on the present study, we have identified the following four heuristics:

1. *Use the plausibility of a hypothesis to choose an experimental strategy.* In this study, we found that both children and adults varied their approach to confirmation and disconfirmation according to the plausibility of the currently-held hypothesis.

2. *Focus on one dimension of an experiment or hypothesis.* An incremental, conservative approach has been found to be effective in both concept attainment (Bruner, et al, 1956) and hypothesis testing (Tschirgi, 1980). This suggests that in moving from one experiment or hypothesis to the next or in moving between experiments and

hypotheses, one should decide upon the most important features of each and focus on just those features.

3. *Maintain observability.* This heuristic depends upon knowledge of one's own information processing limitations as well as knowledge of the device. Our finding that the third graders did not attempt to maintain observability, whereas the sixth graders and adults did, may be a manifestation, in the realm of experimental design, of the more general findings about the development of self-awareness of cognitive limitations (Wellman, 1983).

4. *Design experiments giving characteristic results.* This heuristic maximizes the interpretability of experimental outcomes. Physicians look for "markers" for diseases, and physicists design experiments in which suspected particles will leave "signatures." In the BT domain, this heuristic is instantiated as "use many distinct commands."

Adults not only appeared to use each of these heuristics, but also they appeared to be able to deal with their inherent contradictions. In contrast, children either failed to use these heuristics at all, or else they let one of them dominate.

4. Discovering discovery systems

Because the SDDS framework is applicable to any form of scientific discovery, we can use it to characterize our own endeavors. Most of the effort in the creation of computation models of discovery can be viewed as an attempt to evoke frames in the space of hypotheses stated as running programs. These hypotheses are instantiated as discovery systems, but they are only weakly constrained by empirical evidence from human performance.

However, the H-space search is constrained by other accepted and useful implementations of such theories about discovery as explanation-based learning, abduction, and so on. (See Cheng, 1992; for an illustration of how the different models provide constraints for further search in the space of hypotheses about the discovery process.) This work is based on an implicit normative analysis with the assumptions derived from intuition or logic, rather than from induction over a rich data base.

In contrast, most psychological studies of scientific reasoning can be viewed as search in the space of experiments. Moreover, this E-space search is not usually used for hypothesis testing, but rather it is mainly at the level of either evoking frames or filling slot values. Most of the effort has been focused on empirical studies about the nature of human thinking in situations that approximate "real" scientific discovery.

5. Commonalities between cognitive psychology and machine learning approaches to discovery

Although these two approaches, Cognitive Psychology and Machine Learning, start from quite different points, use different search processes, and use different criteria to evaluate their progress, they are converging on the same general discoveries about the discovery process.

For example, Sleeman, Stacey, Edwards, and Gray (1989) argue that data-driven discovery systems need to be extended in such a way that domain knowledge can be used to constrain the search for the form of the functions that will be tested. This view, derived from a consideration of difficulties encountered in the analytic approach to understanding discovery, is entirely consistent with the results from our empirical approach to the same issues. Indeed, that is what one would hope for; for our basic premise is that search in the two spaces should converge toward discovery. In this case, the entity doing the dual search is the field at large, rather than a single scientist, but I see convergence, nevertheless.

A few things that are wrong, missing, or inadequate in the current work on scientific discovery.

5.1. More spaces

It is clear that scientific discovery takes place in more than only two spaces. Three obvious types of spaces that will be necessary for a full account of scientific discovery include:

1. The *instrumentation space* is clearly a complex and fundamental space in its own right. Machine discovery systems do not worry about this much: they assume that the data are there, waiting to be analyzed by the discovery system, or else they postulate an idealized set of experiments to generate such data.

2. The *representation space* has also received short shrift in my account, although its role is also crucial. Finding the right representation is crucial, and finding it requires constrained search in a large space of possibilities. (Cheng & Simon, 1992; Kaplan & Simon, 1990).

3. The *communication space* includes choices about how to package, disseminate, promote and defend the science, as well as what to read, whom to listen to, what meetings to attend. In many cases, these considerations, of audience, of intended impact, and of how to relate ones work to the existing body of knowledge, have far-reaching impact on the kind of core science that one does (Bazerman, 1988).

5.2. Complexity and knowledge

Some machine discovery systems deal with enormously complex "real world" domains. However, much of the work on discovery—both the construction of machine discovery systems and the psychological studies of discovery—work in highly simplified domains. The question remains about the extent to which our results would scale up when we move to domains in which either prior knowledge or inherent complexity were increased.

5.3. *Social context*

Social exchange provides a rich source of knowledge and constraint in scientific discovery. Sociologists and historians focus almost entirely on processes outside the individual that shape scientific discovery, but they are silent on the cognitive processes that are involved in this social exchange (Bijker, Hughes, & Pinch, 1987). Cognitive psychologists are just beginning to investigate the role of collaboration in scientific reasoning, (Dunbar, 1992).

5.4. *Motivation*

We do not model motivational factors; yet motivation, at the least, must influence the criteria used to accept and reject theories, and it must serve to focus attention on specific goals. Much of the best work in machine discovery is based on the historical record of the great scientists making the great discoveries. But we have been very selective in extracting information from those historical accounts. Such accounts are often filled with statements about excitement, astonishment, disappointment, envy, doubt, and despair. Yet these are not incorporated in any of today's discovery systems. Nor do they get much attention from cognitive psychologists who study scientific discovery.

5.5. *Learning and development*

Why does it take so long to train a scientist? Is it all due to the slow learning rate of humans and the huge amount of content specific knowledge necessary to work in a field? Why don't we start earlier then? Is it because domain-general search constraints are simply not available to young children? It is clear that we have a lot more work to do before we really understand the nature of these cognitive limitations (Glynn, Yeany, & Britton, 1991).

To put the ball in the machine learning court: Machine Discovery, which is a subfield of machine learning, builds systems which are able to make discoveries. But where are the systems that learn how to make discoveries? And finally: What has been discovered in the machine discovery field that has any implications for the teaching of science and the scientific method?

References

Bazerman, C. (1988). *Shaping written knowledge. The genre and activity of the experimental article in science* Madison, WI: University of Wisconsin Press.

Bijker, W. E., Hughes, T.P., & Pinch, T. (1987) *The social construction of technological systems: new directions in the sociology and history of technology* Cambridge, MA: MIT Press

Brewer, W. F., & Samarapungavan, A. (1991). Child theories versus scientific theories: Differences in reasoning or differences in knowledge? In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes Applied and ecological perspectives.* Hillsdale, NJ: Erlbaum.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956) *A study of thinking* New York, NY: Science Editions, Inc.

Cheng, P. C-H (1992). Approaches, models and issues in computational scientific discovery. In M.T. Keane & K. Gilhooly, (Eds.), *Advances in the Psychology of Thinking, Vol. 1* Hemel Hempstead, Hertfordshire: Harvester Wheatsheaf.

Cheng, P. C-H., & Simon, H.A. (1992). The flight representation for discovery: Finding the conservation of momentum. In *Machine Learning. Proceedings of the Ninth International Workshop (ML92)* (pp. 62–71) San Mateo, CA: Morgan Kaufmann.

Dunbar, K., (1989). Scientific reasoning strategies in a simulated molecular genetics environment. In *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society* (pp. 426–433) Hillsdale, NJ: Erlbaum.

Dunbar, K., (1992) Evidence, evaluation and planning heuristics in molecular biology laboratories. *Proceedings of Annual Meeting of Judgement & Decision Making Society*

Dunbar, K., & Schunn, C. D. (1990) The temporal nature of scientific discovery: The roles of priming and analogy. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 93–100). Hillsdale, NJ: Erlbaum.

Gholson, B., Shadish, Jr., W.R., Neimeyer, R.A., & Houts, A.C. (1989) *Psychology of science. Contributions to metascience* Cambridge, MA: Cambridge University Press

Giere, R.N. (1988) *Explaining science. A cognitive approach*. Chicago, IL: University of Chicago Press

Glynn, S.M., Yeany, R.H., & Britton, B.K. (1991). *The psychology of learning science* Hillsdale, NJ: Erlbaum

Gorman, M. E. (1992) Using experiments to determine the heuristic value of falsification. In M. Keane & K. Gilhooly (Eds.) *Advances in the Psychology of Thinking,* Vol. I Hemel Hempstead, Hertfordshire: Harvester Wheatsheaf.

Holland, J., Holyoak, K., Nisbett, R. E., & Thagard, P. (1986) *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press

Kaplan, C.A., & Simon, H.A. (1990) In search of insight. *Cognitive Psychology, 22,* 374–419

Klahr, D., & Dunbar, K. (1988) Dual space search during scientific reasoning. *Cognitive Science, 12,* 1–55

Klahr, D., Fay, A., Dunbar, K. (1993) Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25,* 111-146.

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

Kuhn, D. (1989) Children and adults as intuitive scientists. *Psychological Review, 96,* 674–689

Kulkarni, D., & 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 processes* Cambridge, MA: MIT Press

Qin, Y. & Simon, H.A. (1990). Imagery and problem solving. *Proceedings of The 12th Annual Conference of the Cognitive Science Society* Hillsdale, NJ: Erlbaum

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31–57.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991) Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28,* 859–882

Shrager, J., & Langley, P. (1990) *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann

Simon, H.A. (1966) Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), *Mind and Cosmos* Pittsburgh, PA: University of Pittsburgh Press

Simon, H.A. (1973) Does scientific discovery have a logic? *Philosophy of Science, 40,* 471–480

Sleeman, D.H., Stacey, M.K., Edwards, P., & Gray, N.A.B. (1989) An architecture for theory-driven scientific discovery. In *Proceedings of the European Working Session on Learning (EWSL-89)* (pp. 11–23) London: Pitman

Sodian, B., Zaitchik, D., & Carey, S. (1991) Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62,* 753–766

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51,* 1–10

Tweney, R. D., Doherty, M. E., & Mynatt, C. R. (1981) *On scientific thinking* New York, NY: Columbia University Press.

Valdes-Perez, R.E., Simon, H.A., & Murphy, R.F. (1992) Discovery of pathways in science. In *Proceedings of the Machine Discovery Workshop* at the *International Conference on Machine Learning (ML92)*
Vosniadou, S & Brewer, W F (1992) Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24,* 535–585
Wason, P. C (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140
Wellman, H M (1983) Metamemory revisited. In M. T Chi (Ed.), *Trends in memory development research.* New York, NY: Karger