AI Magazine Volume 8 Number 2 (1987) (© AAAI)

Viewing the History of Science as Compiled Hindsight

Lindley Darden

This article is a written version of an invited talk on artificial intelligence (AI) and the history of science that was presented at the Fifth National Conference on Artificial Intelligence (AAAI-86) in Philadelphia on 13 August 1986 Included is an expanded section on the concept of an abstraction in AI; this section responds to issues that were raised in the discussion which followed the oral presentation The main point here is that the history of science can be used as a source for constructing abstract theory types to aid in solving recurring problem types. Two theory types that aid in forming hypotheses to solve adaptation problems are discussed: selection theories and instructive theories. Providing cases from which to construct theory types is one way in which to view the history of science as "compiled hindsight" and might prove useful to those in AI concerned with scientific knowledge and reasoning

When I was invited to give a talk on AI and the history of science, I declined the suggestion that I discuss the place of AI in the history of science. The field is too new, developing too rapidly, to try to assess its place in the history of science just yet. Some aspects of AI are applied and more like engineering than science. Some aspects, those in which AI researchers aim at modeling human intelligence, merge into cognitive science. Those researchers trying to understand intelligence, human or otherwise, are forging a new area that might or might not ultimately be called scientific. I don't believe it is profitable to argue over whether a field deserves the honorific title of "science." Each field needs to develop its own methodologies, which are shaped by the needs and problems within the field, without succumbing to the disease of physics envy. If some of the many characteristics attributed to science are useful within a new field, then they should be considered. Some fields, for example, might find controlled experimentation useful; others might strive for quantitative results; and still others might find theory-driven methods of use. However, adopting methods simply to obtain the label scientific and not because they are demanded by the problems seems to me to be misguided.

Although I believe it is too early to try to assess AI's place in the history of science, it is not too early to begin preserving records for the historians who will be writing the history of AI. AI is an exciting and rapidly developing field. Those persons shaping it should be aware of the need to preserve records of their activities. Pamela McCorduck's entertaining and journalistic account of AI, *Machines Who Think* (1979,) will be the first of many chronicles of this field. Professional historians will soon follow the journalists.

The difficulty in preserving programs and the hardware on which they run will make the historians task even more difficult than the usual problem, for example, of coping with scientists who cleaned out their files and threw away their notebooks. Those of you working in the field should be careful about preserving material. Consider donating your papers to university or corporate archives, and be aware that future generations will likely want to trace your steps of thinking and implementing. One scientist whose work I have tried to trace cleaned out his files every five years. Another researcher destroyed his notebooks when he retired. Because we cannot date a key discovery on the basis of information in this researcher's notebooks, which are the best historical source, we are left with trying to reconstruct what he said in his classes during this period to see if he discussed as yet unpublished results in his lectures. Would you want to count on your students' lecture notes to substantiate your claim that you made a discovery? Certain large projects, such as the space telescope, hire a historian at the outset so that appropriate records can be kept and interviews done as the project proceeds. Scientists interviewed many years later might or might not accu-

Although I believe it is too early to try to assess AI's place in the history of science, it is not too early to begin preserving records for the historians who will be writing the history of AI.

rately remember what happened and when. Having a historian involved in the early stages is a good way to record AI work on large projects.

The Babbage Institute and the Smithsonian Institution have strong interests in the history of computing, but they have done little yet on the history of AI. Arthur Norberg of the Babbage Institute told me that they will be happy to advise anyone with archival material in AI about appropriate places for storage. A major exhibit on the information revolution is scheduled to open at the Smithsonian Institution in 1990. Currently, nothing on AI is scheduled to appear in this exhibit.

In contrast, the Computer Museum in Boston has an exhibit entitled, "Artificial Intelligence and Robots," opening in the summer of 1987. Oliver Strimpel is the curator working on the exhibit. The museum already has Shakey and other artifacts and is interested in acquiring still more. It is also interested in collecting programs, either in listings or in other usable forms, for their archives. For the exhibit, they need rugged, state-of-the-art AI systems which might be of interest to the public and which could withstand museum use.

The AI field is fortunate to have its key founding fathers still alive. It is certainly time for professional historians--those trained in oral history methodology--to begin collecting audiotapes and videotapes, deciding on key historical questions that need to be asked, and filling in the gaps in documentation. The American Association for Artificial Intelligence might wish to fund such a project, just as many of the scientific societies have done.

Although the history of AI is an interesting topic, it is not the focus of my discussion today. It is not my area of research. I am a historian and a philosopher of the natural sciences (not the social or artificial). I am interested in scientific knowledge and reasoning. My topic deals with the ways in which AI and the history and philosophy of science can profit from interacting. Successes already exist with such an interface, and I would like to suggest additional points of potential contact.

When I was invited to speak, I was asked to be entertaining and provocative. Because I am normally rather matter-of-fact and, I guess, sad to say, a bit dull, such a task isn't especially easy for me. AI has some very bright, clever, and entertaining people who give very lively talks. I can't hope to match their liveliness and wit. Also, being provocative requires one to take dogmatic stands so that others will disagree. Philosophical training has sharpened my ability to see both sides of an issue and dulled my ability to be dogmatic. However, in an effort to satisfy this request, let me attempt to make a few provocative remarks before getting on to the serious--and I hope not too dull--points.

To give an overview of what is coming, first I will make some provocative, introductory remarks. Then, I will discuss fruitful interactions between AI and the history and philosophy of science. Next will come a discussion of an important idea developed in AI, namely, abstraction. This idea will be applied in a discussion of abstractions that I have found by analyzing the history of science as well as an illustration involving two types of theories from the biological sciences--selection and instructive theories. The conclusion is a summary of the compiled hindsight provided by the examples discussed and suggests ways in which AI and the history of science can continue to interact.

AI and Other Fields

In a provocative article in *AI Magazine* in 1983, Roger Schank made some remarks about AI that are worth recalling. Schank discussed the sometimes tenuous connection between AI and the rest of computer science. He said that AI is a methodology applicable to many fields other than computer science because it is "potentially, the algorithmic study of processes in every field of inquiry. As such the future should produce AI anthropologists, AI doctors, AI political scientists, and so on" (Schank 1983, p. 7). I would add to this list AI historians and philosophers of science.

Students educated in computer science as undergraduates who then specialize in AI as graduate students learn a lot of skills, maybe some mathematics, but they might be exposed to very little actual knowledge, scientific or otherwise, or to problems examined by researchers in other fields. I think this lack of training in knowledge-rich fields accounts in part for so many of the toy problems and games that have been the domain of much AI work. AI efforts that address real, rather than toy, problems, especially scientific ones, are of much greater interest to those researchers in other fields attempting to make connections with AI.

The work in expert systems has been all for the good in getting AI researchers involved with real-world knowledge and reasoning. Some AI systems have been driven by research problems and needs in other areas including science. Such success requires those involved in AI to learn about other fields and become concerned with the research problems and reasoning strategies in various content fields. Graduate programs in AI should have less training in computer science. More of the students' time should be spent in research fields where AI techniques can be applied and where the needs of these fields can drive the development of new areas of AI. The history of science has been marked by successful interfield interactions leading to new developments (Darden and Maull 1977). The potential exists for the AI field to be a part of many successful interfield bridges, and as Schank suggested, what would then mark a successful contribution in AI would be developments that are applicable to a wide range of areas (Schank

1983, p.7).

Another case of AI researchers lacking content knowledge but attempting to apply their powerful skills is the current work on commonsense reasoning. As someone working on the relations between the history and philosophy of science and AI, I don't find the work on common sense of particular interest. Common sense is a very vague term. The eclectic list of topics pursued in AI under this rubric demonstrates a lack of clear focus. Furthermore, commonsense notions about the natural world are often wrong. Science has spent hundreds of years cleaning up commonsense misconceptions and improving loose methods of reasoning.

Much of Aristotle's world view was a commonsense one: objects stop moving if no mover is pushing them; the earth is at the center of the universe and does not move, and the stars revolve around the earth; animal species perpetuate themselves eternally and act according to teleological goals; substances have essential and accidental properties. Science, or natural philosophy as it was called until the nineteenth century, has served to correct these and many other commonsense misconceptions. Thus, science is an important storehouse of knowledge about the natural world. Storing this knowledge and finding ways of making it readily accessible to AI systems is an important task. As a philosopher of science, I find efforts to encode and use scientific knowledge of much more interest than the attempts to make the vague concept of commonsense knowledge precise. (For further discussion of the view that epistemology should be the study of scientific, rather than commonsense, knowledge, see Popper 1965, p. 18-23.)

Having now provocatively suggested that AI researchers could profit from working on scientific problems rather than toy ones and dogmatically asserted that science has cleaned up commonsense misconceptions, I will try to turn to claims of greater substance. The history of science and the philosophy of science can be related in various ways (see Figure 1). Much of the history of science now being written is not focused on scientific ideas but on the social and institutional aspects of science. Thus, it is outside the sphere of the philosophy of science. Some work has been done using the social history of science in AI and would, thus, be located in the intersection of the history of science with AI outside the philosophy of science, as shown in figure 1. A possible example is Kornfeld and Hewitt's (1981) effort to build a parallel system that is analogous to a scientific community of researchers that are working independently on the same scientific problem.

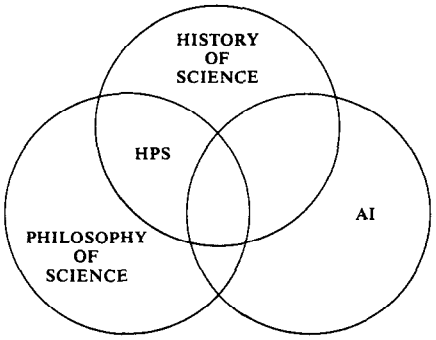


Figure 1: Interfield Relations

The philosophy of science can (and much of it has) involve logical analyses of science without attention to the historical development of scientific ideas. The work of Glymour, Kelly, and Scheines (1983) on implementing the testing of scientific hypotheses in a logical form is an example of philosophers of science (those who aren't concerned with the history of science) making use of AI techniques. Their work would be located in the area of figure 1 in which AI and the philosophy of science intersect outside the history of science.

The history of science and the philosophy of science interact in the interdisciplinary area of the history and philosophy of science (HPS). The aim of HPS is to understand reasoning in the growth of scientific knowledge. HPS researchers make use of extensive case studies from the history of science in order to understand scientific change. I argue that AI and HPS have and can continue to interact to their mutual benefit; important work has been and can continue to be done at the intersection of AI and HPS, the center area in figure 1.

AI and the History of Science

Thus far, I have discussed science as a body of knowledge. However, science can also be characterized by its methodologies, which themselves have improved over time. Some methods used in science have already proved fruitful in AI. Experimentation is often seen as a key method in science. The early work on using AI techniques in molecular genetics (MOLGEN) (Friedland and Kedes 1985) aided in the planning of experiments. One of the most interesting pieces of work in AI from a science historian's point of view is the work by Langley, Bradshaw, and Simon (1983) and others on Baconian induction. They developed methods for finding patterns in numeric data. Another of their systems, GLAUBER, can discover certain kinds of qualitative laws (Langley et al. 1983). These methods are sufficient for rediscovering numerous empirical laws, such as the laws of chemical combination discovered in the nineteenth century. Implementing inductive-reasoning strategies in order to rediscover past scientific results has provided fruitful insights into scientific reasoning methods. But data-driven, bottom-up, inductive methods have their limits, as Langley's group realized (Langley et al. 1986).

Consider a simplistic division of scientific knowledge into data, empirical generalization, and explanatory theory. Inductive methods can be seen as a way of getting from data to empirical generalizations but not to explanatory theories. Explanatory theories require concepts not in the data themselves in order to explain these data (philosophers of science have argued, for example, Hesse 1966). A method for producing explanatory hypotheses is the hypothetico-deductive method: a hypothesis is guessed, and a prediction is then derived from it. The philosopher of science Karl Popper advocated this method and stressed the importance of falsifying instances to eliminate erroneous guesses (Popper 1965).

The hypothetico-deductive method corresponds--or perhaps even gave rise-

to the methodology in AI called "generate and test." The key component becomes the hypothesis generator. The idea from the philosophy of science that generation of hypotheses is by unconstrained guessing is not very helpful in trying to implement a hypothesis generator. The DENDRAL work (Lindsay et al. 1980) was an implementation of a generate-and-test methodology for forming hypotheses within a theoretical context. The hypotheses could be exhaustively specified. However, the hypothesis space was so large that a key problem was to find constraints to add to the generator to restrict the search of the hypothesis space. Use of the generate-and-test method provides a real challenge when the hypothesis space is not well-defined--guessing new theories is not something that computers are good at.

Philosophers of science have criticized the lack of insight provided by the hypothetico-deductive method that hypotheses are to be guessed. N. R. Hanson (1961), drawing on the work of Charles Pierce, argued for an alternative to the hypothetico-deductive method called "retroduction" or "abduction." Abduction is a method of plausible hypothesis formation. It is beginning to receive attention in AI. In their recent AI textbook, Charniak and McDermott discussed abduction in the logical form usually referred to as the "fallacy of affirming the consequent":

If *a*, then *b*
b
Therefore, *a*

Of course, *b* could result from something other than *a*. However, if one already has the reliable generalization that *a* causes *b*, one can plausibly conjecture that if *b* occurred, then *a* would explain its occurrence (Charniak and McDermott 1985, ch. 8). Sophisticated versions of abductive inference are being developed by Reggia, Nau, Wang, and Peng (1985), as well as Thagard and Holyoak (1985).

Hanson's schema for retroduction was more complex than the simple logical form discussed by Charniak and McDermott. Hanson was concerned with the discovery of new general theories, not merely invoking past generalizations in another instance.

Hanson's schema for retroduction was the following: (1) surprising phenomena, p1, p2, p3, . . ., are encountered; (2) p1, p2, p3, . . ., would follow from a hypothesis of H's type; (3) therefore, there is good reason for elaborating a hypothesis of the type of H (Hanson 1961, p. 630).

Hanson had little to say about what constituted a type of hypothesis and gave only a few examples. For instance, he labeled Newton's law of gravitation an inverse-square type of theory. Despite its lack of development, Hanson's idea has appeal: find types of hypotheses proposed in the past, and analyze the nature of the puzzling phenomena to which the type applied. Use this "compiled hindsight" in future instances of theory construction. I think AI can help in the development of Hanson's vague suggestions, for instance, by applying the concept of an abstraction to devising computationally useful theory types.

Two powerful ideas developed by AI researchers in a computationally useful way are abstraction and instantiation.

The Concept of an Abstraction

One of the marks of fruitful interfield interactions in science is that each field contributes to developments in the others. Thus far, I have been indicating how ideas from the history and philosophy of science have been and can be useful in AI. I would also like to suggest how concepts and methods developed in AI might be useful in fully developing methods for hypothesis formation in science.

Two powerful ideas developed by AI researchers in a computationally useful way are abstraction and instantiation. The analysis of analogy as two structures that share a common abstraction is computationally useful in AI (Genesereth 1980; Greiner 1985). An example of a common abstraction is a tree structure; instantiations include the Linnaean hierarchy in biological taxonomy and the organizational chart of a corporation.

The concept of an abstraction has not been well analyzed in AI. Creative new concepts usually start out fuzzy and ill-defined and are used in different ways. Such early vagueness is useful for allowing a new concept to "float" to an appropriate point, for allowing an exploration of its potential areas of application. The use of abstraction is still developing in AI, and this analysis is not meant to be definitive. However, some attempt to analyze the concept of abstraction and to distinguish it from other concepts, such as generalization and simplification, is useful here.

Abstraction can be considered both a process (which will here be called "abstracting") and a product of this process (which will be called an "abstraction"). Abstraction formation involves loss of content. Consider a source and the process of abstracting from it. The abstraction thus produced is, in a sense, simpler than its source. Hence, abstracting is one way of simplifying. Furthermore, because it has less content than the source, an abstraction can apply to a larger class of objects, of which its source is one member. In this sense, abstracting is like generalizing.

The simplifying and generalizing aspects of abstracting are reflected in methods for forming abstractions. Although it is possible to form an abstraction from a single instance, it is easier to consider the case of taking two similar instances and forming the common abstraction of which they are both instantiations. The most straightforward case is to take two instances, delete their differences, and consider their commonalities as the abstraction. With F, G, and H interpreted as properties of an entity a, (F(a) & G(a)) and (F(a) & H(a)) give the abstraction F(a). For example, "the deltoid muscle has actin, and the deltoid muscle has myosin" and "the deltoid muscle has actin, and the deltoid muscle functions as an abductor" abstract to "the deltoid muscle has actin". In this case, content is lost in abstracting: G(a) and H(a)

cannot be retrieved from the abstraction. Thus, the abstraction is simpler than its instantiations.

I don't know if this method of abstracting was developed historically from the famous "axiom of abstraction" in set theory: "given any property, there exists a set whose members are just those entities having that property" (Suppes 1972, p. 6). No specific mention is made of the loss of content here, but the focus on one property to the exclusion of others can involve abstracting from numerous properties to one. This formulation of abstraction in set theory shows why abstraction and generalization are often used synonymously. Polya's work on problem solving has been influential in AI. In his set of definitions, he does not list abstraction but defines generalization in the following way: ". . . passing from the consideration of one object to the consideration of a set containing that object; or passing from the consideration of a restricted set to that of a more comprehensive set containing the restricted one" (Polya 1957, p. 108).

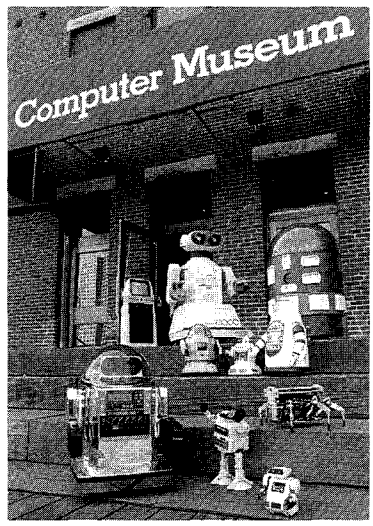
Quine distinguishes two kinds of generality that help explicate the concept of abstraction. The first kind of generality is what he calls the "typically ambiguous," namely, the "schematic generality of standing for any one of a lot of formulas." F(a) is typically ambiguous because numerous substitutions can be made for F and for a. This meaning of generality is close to the usage of abstraction in AI. Second, he discusses the generality of universal quantification, which involves "quantifying undividedly over an exhaustive universe of discourse" (Quine 1969, p. 266). For example, moving from "some muscles contain actin" to "all muscles contain actin" is a step of universal generalization. It would probably not be called abstraction because no content is lost (with the possible exception of the loss of the existential claim that some muscles exist). This example is an instance of inductive generalization, not abstraction.

In forming a general class, however, one can omit individual differences to get the class concept, which involves an abstracting step. For example, consider forming the class concept of cats: "Whiskers has teeth and a white coat," and "Sandy has teeth and a calico coat" are generalized to "all cats have teeth." Generalization (in the sense of universal quantification) occurred in concluding "all cats have teeth," and abstraction occurred in the loss of information about coats (see Harré 1970, for another discussion by a philosopher of science of abstraction formation).

Another method for abstracting produces abstract formulas that are general in the "typically ambiguous" sense. Constants are replaced by variables: F(a) has the abstraction F(x). The changes from (F(a) & G(a)) to F(a) to F(x) involve progressing to higher levels of abstraction. For example, "the deltoid muscle has actin, and the deltoid muscle has myosin" abstracts to "the deltoid muscle has actin", which abstracts to "x has actin". It is easy to see that AI systems would have no difficulty in forming abstractions by dropping conditions or changing constants to variables given appropriate rules for doing so.

In difficult cases, forming an abstraction can involve more than merely dropping parts or replacing constants with variables. New, abstract semantic concepts might have to be introduced. Forming abstractions in such cases might involve creativity in finding the appropriate concepts and terminology. This method might be difficult to implement. Suppose instead of abstracting from "the deltoid muscle has actin" to "x has actin," one wished to abstract to "abductor muscles have actin." The AI system would then have to have an "is a" hierarchy representing the knowledge that the deltoid muscle is an abductor muscle. To move to the next abstract level, the system would substitute the class concept for the individual. Some guidance would be needed to know if the property of the individual (in this case "has actin") applies to the class at the next level in the hierarchy (in this case, it does--abductor muscles do have actin).

The formation of schemas has been discussed by cognitive psychologists. Gick and Holyoak (1983) discuss the example of forming a schema to represent two analogous situations, such as"destroy a tumor by radiation" and "capture a fortress by an army." The schema becomes "overcome a target by a force." "Overcome" is a creative abstraction for "destroy" and "capture." A system that could form such creative abstractions would have to have more semantic content than one which operated merely by replacing constants with variables. The concept of a schematic structure from which specific content has been omitted is often used synonymously with abstraction in AI (see Zeigler and Rada 1984 for a discussion of this sense of abstraction).



Stephen Nelson-Fay Photograph

A robot family portrait taken at the Computer Museum in Boston.

The idea of multiple levels of abstraction is also important in AI. Scientists sometimes make use of mathematical models at one level of abstraction, but multiple levels of abstraction, with each increasingly abstract, is not a common idea in science or the history and philosophy of science. In AI it is easy to consider dropping detail as one proceeds upward in an abstraction hierarchy. Forming the higher-level, abstract concepts while building a knowledge base can be a creative process for the knowledge engineer. The AI researcher might have to add new abstract concepts in order to put the knowledge into a knowledge representation system with multiple levels of abstraction. For example, the concept of a "molecular switch" was developed at a high level of abstraction in the MOLGEN knowledge base, and it has proved of interest

to molecular biologists who hadn't considered their material at such an abstract level (Friedland and Kedes 1985; Karp 1985, personal communication). Progressive refinement by successive instantiations of multiple levels of abstraction is now a common idea in AI, but it is not commonplace in other fields. I will show how it can be a method for producing alternative hypotheses.

Forming the higher-level, abstract concepts while building a knowledge base can be a creative process for the knowledge engineer.

Abstraction has been used in AI systems to do problem solving and planning. Korf (1985) claims that the first explicit use of abstraction in AI was in the planning version of the general problem solver (GPS) developed by Newell and Simon in 1972. A discussion of GPS in the *Handbook of Artificial Intelligence* makes this use clear:

> GPS planned in an abstraction space defined by replacing all logical connectives by a single abstract symbol. The original *problem space* defined four logical connectives, but many problem-solving operators were applicable to any connective. Thus, it could be treated as a detail and abstraction out of the formulation of the problem (Cohen and Feigenbaum 1982, p. 518).

Another early example of the use of abstraction in AI was the work of Sacerdoti (1974) in ABSTRIPS, a planning program.

> ABSTRIPS plans in a hierarchy of abstraction spaces, the highest of which contains a plan devoid of all unimportant details and the lowest of which contains a complete and detailed sequence of problem-solving operators (Cohen and Feigenbaum 1982, p. 517).

For example, in devising a plan to "buy a piano," the highest level might be "locate piano" or "get money." Lower-level details such as "drive to store" would be omitted until the later stages. If the higher-level, critical goals cannot be satisfied, then the lower-level details need never be considered by the planner.

Korf summarizes:

> The value of abstraction is well-known in artificial intelligence. The basic idea is that in order to efficiently solve a complex problem, a problem solver should at first ignore low level details and concentrate on the essential feature of the problem, filling in the details later. The idea readily generalizes to multiple hierarchical levels of abstraction, each focused on a different level of detail. Empirically, the technique has proven to be very effective in reducing the complexity of large problems (Korf 1985, p. 7).

Three methods for forming abstractions from instances have been discussed: (1) dropping conditions, (2) replacing constants with variables, and (3) finding semantically rich concepts at a higher level of abstraction that capture the meaning of the lower-level concepts. What guides the dropping of detail depends on the purpose for which the abstraction is being formed, on the particular problem-solving context. A given entity or even two similar entities might be abstracted in different ways depending on which features are the focus of current investigation (see Utgoff 1986 for a related discussion of bias in learning and Kedar-Cabelli 1985 for a discussion of the role of purpose in forming analogies).

Thus far, the discussion has focused on forming abstractions from instances. However, some AI systems do the opposite; namely, they retrieve available abstractions and instantiate them in a given situation. Friedland's (1979) version of MOLGEN stored skeletal plans containing steps at a high level of abstraction for experiment planning in molecular genetics. In a given problem, the appropriate skeletal plan could be retrieved and instantiated. For example, to splice a gene into a bacterial plasmid, a splicing plan, with its steps of choosing appropriate restriction enzymes and so on, could be retrieved and instantiated for the particular case. This method is similar to Schank and Abelson's (1977) idea of stored scripts to provide expectations in story understanding. Skeletal plans and scripts are stored abstractions based on compiled hindsight.

Thus, to summarize, abstraction formation involves loss of content. The loss makes the abstraction simpler than its instantiation(s). Also, the abstracting process might produce an abstraction that has a larger class of instantiations than the original one(s) from which it was formed; in this sense, abstraction is like generalization. Multiple levels of abstraction can be formed by continuing to apply abstracting methods at each level to produce the next higher one. Several methods for forming abstractions have been discussed, including (1) dropping content; (2) replacing constants with variables; and (3) forming new, abstract semantic concepts. A fourth method of abstraction, which is most important in considering applications from the history of science, is the following: a schema that has resulted from dropping detail and preserving underlying structural or functional relations among component parts.

History of Science as a Source of Abstractions

Except for the development of mathematical models, which can be viewed as going up one level of abstraction, little effort has been made to extract useful abstractions from the history of science that can be instantiated in new problem situations. The thesis to be argued here (philosophers like to argue for theses) is this: *the history of science can serve as a source of compiled hindsight for constructing useful abstractions.* These abstractions will be of theory types or other recurrent patterns and processes in the natural world. They might prove useful in new problem situations--when these situations are of a recurring problem type--for which an available abstraction exists. A presupposition behind this thesis is that recurring patterns and

processes occur in nature. Recurrent types of theories that have proved useful in the history of science are worth analyzing for common features and putting into abstract form. Such abstractions can then serve as compilations of the hindsight gained from their study. Preserving such hindsight becomes especially important when theories are counterintuitive, and scientists have produced insights that serve to correct naive, commonsense views. If this pattern is encountered again, it will be computationally efficient to have an abstraction available rather than to have to rediscover it.

Discussions at the Analogica '85 conference held at Rutgers University, New Brunswick, New Jersey, showed that an important next step in the current work on analogy is to compile instances of useful abstractions. Some of the individuals working on analogy have specifically focused on scientific examples (Darden 1980; Darden 1983). Gentner (1983) discussed central-force systems, such as the solar system and the Bohr model of the atom. Greiner (1985) constructed an abstraction for flow problems, such as water and electric flow. Thagard and Holyoak (1985) investigated wave-type theories, such as water waves and sound waves. The abstractions being developed are computationally useful in doing analogical reasoning and problem solving when appropriate problem types arise. The later stages of the work on Baconian induction were not specifically focused on abstractions in analogical reasoning; nonetheless, they involved reasoning in discovering types of theories: compositional theories, and particulate theories (Langley et al. 1986; Zytkow and Simon 1986).

Bringing a computational AI perspective to case studies in the history and philosophy of science provides a new focus. The task is to find abstract patterns and processes that can be implemented and used in more cases.

Abstractions for Selection and Instructive Type Theories

Bringing this computational perspective to my own area of the history and philosophy of biology has led me to look for abstract characterizations of biological theories and the reasoning that could produce them. (Although I have begun trying to implement reasoning sufficient for rediscovering a theory in genetics, this implementation is very much in the formative stages. Its discussion is appropriately left for technical sessions in future years (see Darden and Rada forthcoming, for a preliminary sketch). I will now discuss an example of a problem type and a recurrent theory type that has resulted from searching for useful abstractions which can perhaps be implemented.

An adaptation problem involves explaining how something comes to fit something else, or how two things change over time so that one comes to be adapted to the other. Two different theory types have been proposed historically to solve adaptation problems: selection theories and instructive theories.

An abstraction of selection theories has the following components: variation, interaction of the variants in a constrained environment, and the resultant perpetuation of the fit variants (reproduction or amplification).

This theory type has only been available since the middle of the nineteenth century when Darwin proposed his theory of natural selection to explain the origin of new, adapted species. Darwin amassed evidence to show that variation occurs in nature. He argued that the environment cannot sustain all the organisms which are produced, so there is a struggle for existence. The adapted variants tend to survive. In natural selection, survival into the next generation by reproduction is what ultimately counts for a variant's success. Reproduction can be included at a lower level of abstraction as one of several ways to instantiate the perpetuation of the fit variants. By removing it from the higher-level abstraction for selection theories, the abstraction applies to additional selection theories that don't have a step for reproduction. Once Darwin formed his theory of natural selection, selection as a type of theory became available. It was a product of hundreds of years of biological thought and represented a departure from commonsense observations so radical that some still resist this theory.

Once Darwin formulated a selection theory to explain species change, the theory type was available for other instances of theory construction. An example comes from immunology. Antibodies that are adapted to eliminating invaders in the body need to be generated; thus, formulating a theory of antibody production involves solving an adaptation problem. In improving a theory for antibody formation in the 1950s, Burnet specifically used an analogy to natural selection. He proposed the clonal selection theory: cells vary according to the antibodies they produce, and a large amount of diversity is present in the naturally circulating antibodies. Selective interaction with invaders occurs. Those cells which produce the appropriate antibodies reproduce in clones and produce an amplified level of the particular antibody until the invader is eliminated (if all goes well). Both Darwinian natural selection and clonal selection have fared well in subsequent tests and with further developments in their fields.

The history of science can serve as a source of compiled hindsight for constructing useful abstractions.

Selection as a type of theory can now be seen as a historically successful means of biological theory construction for adaptation problems. It can now be used as an abstract pattern for forming new theories to solve new adaptation problems.

An interesting attempt to formulate a new selection theory is Edelman's "group-selective theory of higher brain function" (Edelman and Mountcastle 1978). As far as I know, it has not yet been confirmed or disconfirmed. The unit of variation is a group of brain cells. Groups of cells interact with sig-

nals such that some are amplified and others are not. This amplification can occur by either positive or negative selection of the groups of cells. Because the theory is still in its formative stage, alternative ways of instantiating the abstraction provide competing hypotheses to be tested, such as this case of positive or negative selection. As other theories are formed that instantiate the abstraction, then alternatives at a slightly lower level of abstraction can be elaborated to provide alternative hypotheses. It will be interesting to see the fate of this new selection theory.

Another type of theory predated selection theories as an alternative for solving adaptation problems--instructive theories. Instead of a random set of variants, information is used to generate variants that are already fit. Neo-Lamarckian inheritance of acquired characteristics is an example. A new form is generated in a changed environment to be adapted to the constraints in the environment. For example, the giraffe sees the leaves high in the tree and grows a longer neck to reach them, which is then inherited by the baby giraffe. No wasted variants are produced, and no selection is necessary to produce the adaptation. Although this theory eliminates the waste of producing unadapted variants, it necessitates a powerful generator of new forms. The generator has to be able to receive information from the changing environment and make new forms that are adapted to these changes.

Historically, an instructive theory was also proposed in immunology. The template theory of antibody formation (Pauling 1940) proposed that amorphous preantibodies were generated. An invader was then used as a template and the preantibody formed into an antibody by wrapping around it. Neither of the instructive theories was confirmed historically. In both biological evolution and antibody production, the selection theories have proved the successful types. Hindsight thus shows, as Edelman succinctly said, "It is clear from both evolutionary and immunological theory . . . that in facing an unknown future, the fundamental requirement for successful adaptation is preexisting diversity" (Edelman and Mountcastle 1978, p. 56).

However, in a known future--namely, a stable environment--developing a generator to make just adapted forms might well be the better strategy. One might, for instance, apply these types of theories to the adaptation problem of making a part in a factory that is adapted to another part. An instructive process would be better than a selective one if the kind of part to be produced will be stable over a sufficiently long period of time, and a means exists for communicating to the mechanism generating the part the specifications the part should meet. In another example, debates in evolutionary epistemology have raised the question of whether humans blindly generate alternatives when forming new concepts and then select them or whether a directed, instructive process better captures human knowledge acquisition (see Campbell 1974 and Thagard forthcoming for discussions of evolutionary epistemology).

In considering the hindsight provided by these historical cases, consideration of the incorrect theories, as well as the correct ones, has proved useful. Scientists often consider disproved theories in a scientific graveyard, not worth a second glance by those researchers pushing ahead at the forefront. However, in these cases, incorrect but plausible, as well as confirmed, theories proved worth considering. The disconfirmed theories provided a theory type that might be useful in cases outside the biological examples from which the type was abstracted. Also, understanding the reason the disconfirmed theories failed in these biological cases provided additional hindsight about appropriate conditions for instantiating one or the other of the two types.

Conclusion

The compiled hindsight from the study of selection and instructive theories is the following: find a current problem in which something is to change over time to fit something else; devise good abstractions of the theory types for adaptation problems, namely, selection and instructive theories; develop criteria for choosing which abstraction to instantiate; instantiate the appropriate abstraction in the current adaptation problem situation to provide new hypotheses; test the hypotheses; and refine the criteria for applying the abstraction in light of successes and failures.

A general research program emerges from the thesis that the history of science provides compiled hindsight: study the history of science to find recurring problem types and theory types, devise computationally useful abstractions for them, and build AI systems to use such compiled hindsight in new problem situations.

Important tasks in the coming years will be to put scientific knowledge into a form that can be computationally useful and to devise reasoning strategies which can be implemented to produce and make use of scientific knowledge. Successful use of the compiled hindsight from the history of science could contribute to this task and show the usefulness of interfield interactions between the history of science and AI.

Earlier at the conference, I had lunch with Doug Lenat, from whom I have taken the phrase "compiled hindsight" (Lenat 1983). I complimented Doug on his ability to give very entertaining talks. He said one rule is to close with a joke; then people go away thinking they enjoyed your talk. However, I don't have any good abstractions for good jokes in AI, so this is one piece of compiled hindsight that won't be instantiated here.

Acknowledgments

Lindley Darden gratefully acknowledges the support of the University of Maryland Institute for Advanced Computer Studies and the College of Arts and Humanities, University of Maryland, for this work. Thanks to Joseph Cain, Subbarao Kambhampati, Joel Hagen, Pamela Henson, Roy Rada, and colleagues at the National Library of Medicine for comments on an earlier version of this talk. Thanks to Mark Weiser, James Platt, David Itkin, and Gwen Nelson for help in getting the manuscript into final form.

References

Burnet, F. M 1957 A Modification of Jerne's Theory of Antibody Production

Using the Concept of Clonal Selection. *The Australian Journal of Science* 20:67-69.

Campbell, D. T 1974 Unjustified Variation and Selective Retention in Scientific Discovery In *Studies in the Philosophy of Biology,* eds. F J Ayala and T Dobzhansky, 139-161. Berkeley: University of California Press

Charniak, E , and McDermott, D. 1985. *Introduction to Artificial Intelligence* Reading, Mass.: Addison-Wesley.

Cohen, P., and Feigenbaum, E., eds. 1982. *Handbook of Artificial Intelligence,* vol. 3. Los Altos, Calif.: Kaufmann.

Darden, L. 1983 Reasoning by Analogy in Scientific Theory Construction. In Proceedings of the International Machine Learning Workshop, ed. R. Michalski, 32-40. Urbana-Champaign, Ill : Univ. of Illinois.

Darden, L. 1980. Theory Construction in Genetics. In *Scientific Discovery: Case Studies,* ed. T. Nickles, 151-170. Dordrecht, Netherlands: Reidel

Darden, L., and Maull, N. 1977. Interfield Theories. *Philosophy of Science* 44:43-64.

Darden, L., and Rada, R Forthcoming. Hypothesis Formation via Interrelations. *Analogica: The First Workshop on Analogical Reasoning,* ed. A. Prieditis. London: Pitman.

Edelman, G , and Mountcastle, Y. 1978. *The Mindful Brain, Cortical Organization and the Group Selective Theory of Higher Brain Function* Cambridge, Mass.: MIT Press.

Friedland, P. 1979. Knowledge-Based Experiment Design in Molecular Genetics, Technical Report, 79-771, Dept. of Computer Science, Stanford Univ.

Friedland, P., and Kedes, L. 1985 Discovering the Secrets of DNA. *Communications of the ACM* 28:1164-1186

Genesereth, M 1980. Metaphors and Models. In Proceedings of the First Annual National Conference on Artificial Intelligence, 208-211. Menlo Park, Calif.: American Association for Artificial Intelligence.

Gentner, D. 1983. Structure Mapping--A Theoretical Framework for Analogy. *Cognitive Science* 7:155-170

Gick, M , and Holyoak K. 1983. Schema Induction and Analogical Transfer. *Cognitive Psychology* 15:1-38.

Glymour, C.; Kelly, K.; and Scheines, R. 1983 Two Programs for Testing Hypotheses of Any Logical Form. In Proceedings of the International Machine Learning Workshop, ed. R. Michalski, 96-98. Urbana-Champaign, Ill.: Univ. of Illinois.

Greiner, R. 1985 Learning by Understanding Analogies, Technical Report, STAN-CS-85-1071, Dept. of Computer Science, Stanford Univ.

Hanson, N R 1961 (1970). Is There a Logic of Scientific Discovery? In *Current Issues in the Philosophy of Science,* eds. H. Feigl and G. Maxwell New York: Holt, Rinehart, and Winston. Reprint. In *Readings in the Philosophy of Science,* ed. B. Brody, 620-637. Englewood Cliffs, N J.: Prentice-Hall

Harré , R. 1970 *The Principles of Scientific Thinking.* Chicago: University of Chicago Press.

Hesse, M. 1966. *Models and Analogies in Science* Notre Dame, Ind.: University of Notre Dame Press.

Kedar-Cabelli, S. 1985. Purpose-Directed Analogy In Proceedings of the Seventh Annual Conference of the Cognitive Science Society, 150-159. Irvine, Calif.

Korf, R. E. 1985. An Analysis of Abstraction in Problem Solving. In Proceedings of the 24th Annual Technical Symposium, 7-9. Gaithersburg, Md : Washington, D.C., Chapter of the Association for Computing Machinery

Kornfeld, W., and Hewitt, C. 1981. The Scientific Community Metaphor *IEEE Transactions on Systems, Man, and Cybernetics* SMC-11: 24-33.

Langley, P ; Bradshaw, G ; and Simon, H. 1983. Rediscovering Chemistry with the BACON System. In *Machine Learning,* eds. R. Michalski, J. Carbonell, and T. Mitchell, 307-329. Palo Alto, Calif.: Tioga.

Langley, P.; Zytkow, J.; Bradshaw, G.; and Simon, H. 1983 Three Facets of Scientific Discovery. In Proceedings of the Eighth International Joint Conference on Artificial Intelligence, vol. 1, 465-468. Menlo Park, Calif.: American Association for Artificial Intelligence.

Langley, P.; Zytkow, J.; Simon, H.; and Bradshaw, G. 1986. The Search for Regularity: Four Aspects of Scientific Discovery. In *Machine Learning,* vol. 2, eds. R. Michalski, J Carbonell, and T Mitchell, 425-469. Los Altos, Calif.: Morgan-Kaufmann.

Lenat, D. 1983. The Role of Heuristics in Learning by Discovery: Three Case Studies In *Machine Learning,* eds. R. Michalski, J. Carbonell, and T. Mitchell, 243-306 Palo Alto, Calif.: Tioga.

Lindsay, R.; Buchanan, B G.; Feigenbaum, E. A.; and Lederberg, J. 1980. *Applications of Artificial Intelligence for Organic Chemistry, The DENDRAL Project* New York: McGraw-Hill.

McCorduck, P 1979. *Machines Who Think* San Francisco: Freeman.

Pauling, L. 1940 A Theory of the Formation of Antibodies. *Journal of the American Chemical Society* 62:2643-2657.

Polya, G. 1957. *How to Solve It* Garden City, N.Y. Doubleday.

Popper, K. 1965 *The Logic of Scientific Discovery.* New York: Harper Torchbooks

Quine, W. O 1969. *Set Theory and Its Logic* Cambridge, Mass : Harvard University Press

Reggia, J.; Dana, N ; Wang, P.; and Peng, Y. 1985. A Formal Model of Diagnostic Inference. *Information Sciences* 37:227-285.

Sacerdoti, E. 1974 Planning in a Hierarchy of Abstraction Spaces. *Artificial Intelligence* 5:115-135.

Schank, R. 1983 The Current State of AI: One Man's Opinion. *AI Magazine* 4(1):3-8.

Schank, R., and Abelson, R. P. 1977. *Scripts, Plans, Goals, and Understanding.* Hillsdale, N.J.: Lawrence Erlbaum.

Suppes, P. 1972. *Axiomatic Set Theory.* New York: Dover.

Thagard, P. Forthcoming. *Computational Philosophy of Science.* Cambridge, Mass.: MIT Press.

Thagard, P., and Holyoak, K 1985. Discovering the Wave Theory of Sound: Inductive Inference in the Context of Problem Solving. In Proceedings of the Ninth International Joint Conference on Artificial Intelligence, 610-612. Menlo Park, Calif.: American Association for Artificial Intelligence.

Utgoff, P. 1986 Shift of Bias for Inductive Concept Learning. In *Machine Learning,* vol. 2, eds R. Michalski, J. Carbonell, and T. Mitchell, 107-148. Los Altos, Calif : Morgan-Kaufmann.

Zeigler, B., and Rada, R. 1984. Abstraction in Methodology: A Framework for Computer Support *Information Processing and Management* 20:63-79

Zytkow, J., and Simon, H. 1986. A Theory of Historical Discovery: A Construction of Componential Models. *Machine Learning* 1:107-136.