Title

Abduction and World Model Revision

Permalink

https://escholarship.org/uc/item/0zr2z48x

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 11(0)

Authors

O'Rorke, Paul
Morris, Stephen
Schulenburg, David

Publication Date

1989

ABDUCTION AND WORLD MODEL REVISION [1]

Paul O'Rorke
University of California, Irvine

Steven Morris
University of California, Irvine

David Schulenburg
University of California, Irvine

ABSTRACT

Abduction is the process of constructing explanations. This paper suggests that abduction is a key to "world model revisions" — dramatic changes in systems of beliefs such as occur in children's cognitive development and in scientific revolutions. The paper describes a model of belief revision based upon *hypothesis formation by abduction.* When a contradiction between an observation and an existing model or theory about the physical world is encountered, the best course is often simply to suppress parts of the original theory thrown into question by the contradiction and to derive an explanation of the anomalous observation based on relatively solid, basic principles. This process of looking for explanations of unexpected new phenomena can lead by abductive inference to new hypotheses that can form crucial parts of a revised theory. As an illustration, the paper shows how one of Lavoisier's key insights during the Chemical Revolution can be viewed as an example of hypothesis formation by abduction.

BELIEF REVISION USING HYPOTHESES FORMED BY ABDUCTION

"World model revision" is at the more difficult, more creative end of the spectrum of belief revision problems. We all make simple changes in beliefs during everyday life, but dramatic changes in systems of beliefs such as occur in scientific revolutions appear to require extraordinary creative genius. Great changes in our way of looking at the world represent the height of human intellectual achievement and are identified with intellectual giants such as Galileo, Newton, Lavoisier, and Einstein.

James Bryant Conant argues in his introduction to the *Harvard case histories in experimental science* (Conant, Nash, Roller, & Roller, (Eds.), 1957) that case studies of revolutionary advances in science can facilitate the understanding of science by non-scientists. Cognitive scientists take this one step further and argue that case studies based on the history of science can be used to achieve a deeper understanding of the cognitive processes underlying scientific discovery (see, e.g., Bradshaw, Langley, & Simon, 1983; Langley, Simon, Bradshaw, & Zytkow, 1987). One immediate aim of such case studies of scientific revolutions is to develop computational models of the evolution of specific scientific theories over time. However, the ultimate goal is not so much to capture individual case histories — the main goal is to improve our understanding of how theory shifts are, or can be, made.

---

[1] This paper is based on work supported by an Irvine Faculty Fellowship from the University of California, Irvine Academic Senate Committee on Research and by grant number IRI-8813048 from the National Science Foundation to the first author.

The claim of this paper is that *hypothesis formation by abduction* can play a crucial role in world model revision. Abduction is the process of constructing explanations (Peirce, 1931-1958; Pople, 1973; Reggia & Nau, 1984; Schank, 1986; Charniak, 1988; Josephson, Chandrasekaran, Smith, & Tanner, in press). This paper focuses on abduction as a theory driven process. If a prediction of a given theory contradicts an observation, and if methods exist for identifying questionable details of the given theory, this form of abduction can be used to derive an explanation of the anomalous observation based on relatively solid, basic principles of the domain. The claim is that the process of looking for explanations of unexpected new phenomena can lead by abductive inference to new hypotheses that can form crucial parts of new theories.

THE CHEMICAL REVOLUTION

As an illustration, we present some initial results of a case study of the Chemical Revolution — the replacement of the phlogiston theory by the oxygen theory. This particular theory shift has attracted a great deal of interest partly because it occurred in the early days of chemistry, while the theory and the experiments were still close to common knowledge and everyday experience, and were not too highly technical. In addition, the Chemical Revolution has the advantage that a great deal is known about it, because of detailed records left by the scientists involved and due to the large number of books and papers on the subject by historians and philosophers of science (see, for example, Guerlac, 1961; Thagard in press-a; Ihde, 1980; in addition to Conant, 1957).

Prior to the Chemical Revolution, the phlogiston theory of chemistry provided the predominant explanation of the processes of combustion and calcination. Under this theory developed by the German chemist G. E. Stahl (1660 1734), it was thought that all combustible substances contained an element called phlogiston. Combustion was thought of as a sort of flow of phlogiston from combustible substances into the surrounding air. Calcination (e.g., rusting) was also thought of as a loss of phlogiston from metals and metallic calxes[2]. Lavoisier, the 18th century French chemist who was the driving force behind the Chemical Revolution, placed great importance on the observation that the weights of some substances increase in combustion and calcination. Just after this augmentation effect was demonstrated conclusively by experiments, Laviosier deposited a sealed note on November 1, 1772 with the Secretary of the French Academy of Sciences:

> *About eight days ago I discovered that sulfur in burning, far from losing weight, on the contrary, gains it; it is the same with phosphorus... This discovery, which I have established by experiments, that I regard as decisive, has led me to think that what is observed in the combustion of sulfur and phosphorus may well take place in the case of all substances that gain in weight by combustion and calcination; and I am persuaded that the increase in weight of metallic calxes is due to the same cause.*[3]

Lavoisier went on to discover that — contrary to the century old phlogiston theory — a gas contained in the atmosphere combines with burning combustibles and calcinating metals. This gas was first isolated by heating mercurius calcinatus (red calx of mercury; now called red oxide of mercury) until the gas in the calx was liberated. Lavoisier named the new gas "oxygen."

In the next section, we show how advances in research on qualitative physics provide a language for describing some important ideas associated with the phlogiston theory of combustion.

---

[2] A calx is a substance produced by calcination.

[3] Translation by Conant (Conant, 1957). The dots indicate text omitted by the authors.

Direct Influences:
GL1: deriv-sign(Q1, Sign) ← process(Process), active(Process), influence(Process, Q1, Sign).
Indirect Influences:
GL2a: deriv-sign(Q1, Sign) ← qprop(Q1, Q2, pos), deriv-sign(Q2, Sign).
GL2b: deriv-sign(Q1, Sign1) ← qprop(Q1, Q2, neg), deriv-sign(Q2, Sign2), opposite(Sign1, Sign2).
The Law of Sums:
GL3: qprop(Q, Q_i, pos) ← qty-eq(Q, qty-sum(Qs)), member(Q_i, Qs).
The weight of an object is qualitatively proportional to the amount.
GL4: qprop(weight(P), amount(P), pos).
~~**Combustion is a negative influence on the amount of phlogiston in charcoal.**~~
~~GL5a: influence(combustion, amount-of-in(phlogiston, charcoal), neg).~~
~~**Calcination is a negative influence on the phlogiston in mercurius calcinatus.**~~
~~GL5b: influence(calcination, amount-of-in(phlogiston, m-c), neg).~~

The amount of a complex substance equals the sum of the amounts of the components.
GL6: qty-eq(amount(C), qty-sum(Qs)) ← complex(C), is-a-set-of-amounts-of-components-of(Qs, C)
GL7a: is-a-set-of-amounts-of-components-of([Qi | Qs], C) ← is-an-amount-of-a-component-of(Qi, C), is-a-set-of-amounts-of-components-of(Qs, C).
GL7b: is-a-set-of-amounts-of-components-of([ ], C).
GL8: is-an-amount-of-a-component-of(Qi, C) ← complex(C), component(Ci, C), Qi = amount-of-in(Ci, C).

Observation: The weight of mercurius calcinatus increases.
O1: deriv-sign(weight(m-c), pos).

Case facts: Calcination is an active process.
CF1: process(calcination).
CF2: active(calcination).

Figure 1: A Fragment of a Phlogiston Theory, An Observation, and Some Case Facts

SOME ASPECTS OF THE PHLOGISTON THEORY ENCODED AS RULES

Figure 1 shows a fragment of the phlogiston theory describing the effects of combustion and calcination coded in terms of facts and rules. (Ignore the black lines in Figure 1 for now.) Also shown is an observation O1 which describes an increase in weight of a partially calcinated piece of mercury, so-called mercurius calcinatus (here abbreviated m-c). Additionally, case facts CF1 and CF2 indicate that calcination is taking place in some specific situation. This theory, observation and case facts, are expressed in a language derived from Ken Forbus's *Qualitative Process Theory* (Forbus, 1984). In the remainder of this section we briefly describe the individual statements in the fragment of the phlogiston theory.

In Figure 1, rules GL1 and GL2 are general laws of QP theory. GL1, *The Law of Direct Influences*, states that a quantity may be changing because some process is directly influencing it. The quantity increases or decreases according to whether Sign is "positive" or "negative." In this law, "deriv-sign(Q1, Sign)" means "the sign of the derivative of quantity Q1 is Sign".

GL2a and GL2b, *The Laws of Indirect Influences*, are meant to capture the notion that a quantity may change because it is qualitatively proportional to some other quantity. Here "qprop(Q1, Q2, pos)" means "quantity Q1 is positively qualitatively proportional to the quantity Q2." A qualitative proportionality may be either positive or negative. A change in one quantity may be accounted for by a similar change in some other quantity if there is a positive qualitative proportionality between them. In the case of a negative qualitative proportionality, a change in one quantity may be accounted for by an opposite change in another quantity.

Rules GL3, GL4, and GL5 are meant to capture some important aspects of the phlogiston theory. GL3, *The Law of Sums*, states that a quantity is qualitatively proportional to a second quantity if the first quantity is equal to a sum of a number of quantities one of which is the second quantity. "qty-eq(Q, qty-sum(Qs))" means "Q is a quantity equal to the sum of quantities Qs," where Qs is a list of quantities. "member(Qi, Qs)" means "Qi is a member of the list of Qs." GL4 states that the weight of any substance is proportional to the amount of the substance.

Phlogiston theorists viewed all combustible substances as complex substances containing phlogiston. In our qualitative process description of the phlogiston theory, rule GL5a states that combustion is a process that influences the amount of phlogiston in charcoal negatively. That is, if combustion is active, it drives down the amount of phlogiston in a partially burned piece of charcoal. Similarly, rule GL5b states that calcination drives down the amount of phlogiston in a partially calcinated piece of mercury. According to the phlogiston theory, pure metallic calxes were more primitive substances than metals. Metals were formed by heating calxes in the presence of a source of phlogiston such as charcoal; the calxes combined with the phlogiston to form the metals. On the other hand, metallic calxes resulted when phlogiston, which was viewed as a "metallizing principle," flowed out of metals.

Rules GL6, GL7 and GL8 provide some facts about complex substances. These rules state that the amount of a complex substance is equal to the sum of the amounts of its components.

ABDUCTION OF ASPECTS OF THE OXYGEN THEORY

In this section we show how the facts and rules in Figure 1 can be used to construct explanations of observations involving changes in the weights of burning and calcinating substances. In particular, we illustrate the role of abduction in theory formation by showing how Lavoisier's insight can be seen as abductive inference. This is done by showing how a specific "abduction engine", called AbE, generates an explanation of the increase in the weight of calcinating mercury. AbE is a PROLOG meta-interpreter that constructs explanation trees, evaluates partial explanations, and uses best-first heuristic search.

Let us assume as *given* the phlogiston theory shown in Figure 1. The phlogiston theory explains and predicts a *decrease* in the weight of substances undergoing combustion or calcination. This prediction contradicts the given observation that the weight of mercurius calcinatus *increases* during calcination. Assume that, as a result, questionable parts of the theory responsible for the contradiction have been identified and deleted as indicated by the black lines through offending statements in Figure 1.[4] Assume then, that our abduction engine AbE is given the reduced phlogiston theory and the observation and case facts shown in Figure 1. In the reduced theory, phlogiston is no longer considered to be an essential component of combustible substances and no mention is made of the effects of combustion or calcination on amounts of phlogiston.

---

[4] Existing contradiction backtracing and truth maintenance methods could contribute to identifying candidates for deletion or temporary suppression, but some method of evaluating plausibility will be needed in order to decide that a candidate should be suppressed. Basic principles which contribute to many explanations (e.g., conservation laws), should be preferentially retained.

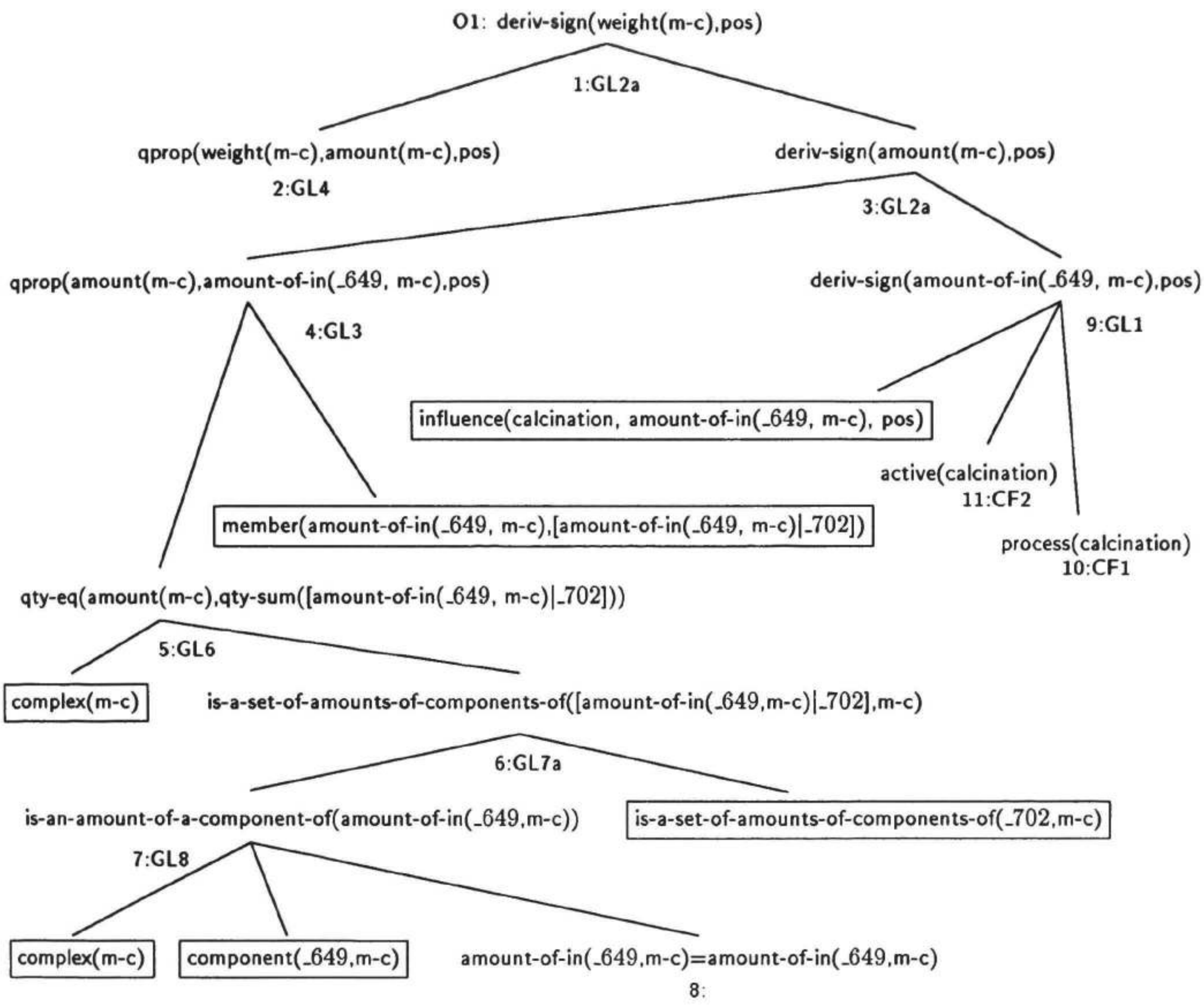

Figure 2: Why the weight of mercurius calcinatus (m-c) increases in calcination.

AbE is asked to explain, in terms of the given laws of qualitative physics and the ablated phlogiston theory, the observation that, during calcination (CF1 & CF2), the weight of mercurius calcinatus increases (O1). AbE does this by attempting to reduce the observation to the given facts, but if this is not possible it will propose some hypotheses in an effort to explain the observation. Figure 2 shows one explanation arrived at by AbE. We now briefly describe how this explanation tree was constructed.

The initial query is: *Why is the weight of the mercurius calcinatus increasing?* According to the laws of indirect influences (GL2), a change in some quantity may be explained by a change in some other quantity provided the two quantities are qualitatively proportional. Backward chaining on this law, AbE proposes that the weight of the mercurius calcinatus may be positively qualitatively proportional to another quantity. The question of whether there is any such quantity is answered as an instance of the general fact that the weight of any object is positively proportional to the amount of that object (GL4). The initial query can thus be explained in terms of an increase in the *amount* of the mercurius calcinatus.

Why is the amount of mercurius calcinatus increasing? To explain this, AbE again uses GL2a to propose a positive qualitative proportionality between the amount of mercurius calcinatus and some other increasing quantity. An appropriate proportionality is found using the law of sums (GL3). Recall that this law states that some quantity Q is proportional to some other quantity Qi if Q is equal to the sum of some set of quantities Qs and Qi is a member of that set. In this case, Q is the amount of the mercurius calcinatus.

The question is whether there is some set of quantities whose sum is equal to the amount of the mercurius calcinatus. This question is answered in terms of knowledge about complex substances (GL6, GL7, GL8). AbE backward chains on these laws to hypothesize that the amount of mercurius calcinatus is increasing because it is a complex substance and the amount of one of its components is increasing. AbE hypothesizes the existence of an unknown quantity of an unknown component of mercurius calcinatus. AbE also hypothesizes a set of remaining components and quantities, without identifying any particular elements of this set.

The question now is whether the amount of the unknown component of mercurius calcinatus is increasing. The law of direct influences (GL1) can be used to explain this increase, assuming that an active process can be found to have a *positive* influence on the amount of the component of the mercurius calcinatus. At this point, since calcination is known to be an active process, AbE completes its explanation by hypothesizing that calcination is a direct positive influence on the amount of the unknown component.

The hypotheses generated by abductive inferences made by AbE during its construction of this explanation of the augmentation of mercurius calcinatus are enclosed in boxes in Figure 2. These abductive inferences correspond to Lavoisier's insight that something was being added during calcination.

RELATION TO OTHER WORK IN COGNITIVE SCIENCE

This work is part of a coherent program of research on automated abduction and machine learning underway at Irvine. Our goal is to explore domain-independent models of abduction and learning in the context of specific examples and domains involving *logical*, *physical*, and *psychological* explanations. Our previous work on logical explanations includes experimental work on explanation-based learning in logical domains such as Principia Mathematica (O'Rorke, 1987). In collaboration with Andrew Ortony and Gerald DeJong of Illinois, we are investigating psychological explanations involving emotions. Initial progress on this research has been reported in (O'Rorke & Cain, 1988). The present paper describes initial progress of our work involving physical explanations. It fits into the theoretical framework for learning in physical domains sketched in Forbus and Gentner (1986). The learning taking place in our chemical revolution example appears to fit in the third stage ("learning naive physics") of Forbus and Gentner's four stage model.

Recent scientific discovery work by Jan Zytkow and Herbert Simon, followed up by Don Rose and Pat Langley, resulted in systems that can automatically detect and correct errors in chemical theories. These artificial intelligence programs, STAHL (Zytkow & Simon, 1986) and STAHLp (Rose & Langley, 1986) are similar, in that they both represent chemical theories in terms of *reaction and component models.* These systems could conceivably model the shift from the phlogiston to the oxygen theory as a change from a set of reaction rules and component models involving phlogiston to a set of reaction rules and component models involving oxygen. In our opinion, however, such an account of the theory shift would be incomplete; if only because the models of the phlogiston and oxygen theories would be incomplete if limited to reactions and component models. For example, both the phlogiston theory and the oxygen theory explained why a flame burning in an enclosed place eventually expires — but these explanations cannot be expressed in terms of component models and reactions alone.

While some revisions in STAHLp amount to hypothesizing the existence of unobserved substances in the input reactions (adding substances), and retracting previously believed observations of substances in the input reactions (deleting substances), all such substances must have been named in previous input reactions. STAHLp is not capable of hypothesizing the existence of a new substance — one that has not previously appeared in an input to STAHLp. This is in contrast to the example we have presented, in which a new component substance is hypothesized on the basis of general qualitative physical laws.

Paul Thagard has also done closely related research. Thagard (in press-b) presents a theory of explanatory coherence and a connectionist implementation. His program, ECHO, is given data representing observations and the phlogiston and oxygen theories. Using activation and inhibition links between data and theoretical statements, the program attempts to determine which of the two theories best "coheres" with the data. Thagard's ECHO focuses on the *evaluation* of *existing* theories. In another paper, (Thagard in press-a) he looks at the conceptual changes that occurred during the overthrow of the phlogiston theory, and gives a fairly detailed conceptual map of several important intermediate stages of chemical theory in the transition from the phlogiston theory to the oxygen theory. In this paper, Thagard suggests that the mechanisms for concept formation and rule abduction present in a program called PI can be used to form conceptual networks that can chart the conceptual changes which occurred during the Chemical Revolution. Our contribution is that we have shown a detailed example of how abduction can be used in concert with ideas from work on qualitative physics to make some crucial inferences associated with the discovery of oxygen.

CONCLUSION

Theory revision can profitably be viewed as a process that involves hypothesis formation by abduction. When an anomaly is encountered, the best course is often simply to forget or suppress questionable details of the original theory and to derive an explanation of the anomalous observation based on more solid, more basic principles. In this way, the process of looking for explanations of unexpected new phenomena can lead by abductive inference to new hypotheses that can form crucial parts of a revised theory.

The main result of this paper is that recent progress on abduction and qualitative process theory makes it possible to automate significant aspects of the reasoning that occurred in the Chemical Revolution. We believe that the language for describing processes and causal relationships resulting from work on qualitative physics together with inference mechanisms such as automated abduction will enable automation of many crucial but relatively common-sense insights associated with scientific revolutions. If this proves true, it suggests that automated abduction is a key to understanding "world model revision."

ACKNOWLEDGEMENTS

Special thanks are due to Pat Langley and Deepak Kulkarni for numerous discussions of scientific discovery. Discussions with Pat Langley and Don Rose on the STAHLp program sparked our interest in modelling the Chemical Revolution. Thanks also to Paul Thagard, of the Princeton University Cognitive Science Laboratory, for several useful discussions of the Chemical Revolution and of scientific revolutions in general.

REFERENCES

Bradshaw, G. F., Langley, P. W., & Simon, H. A. (1983). Studying scientific discovery by computer simulation. *Science, 222*, 971–975.

Charniak, E. (1988). Motivation analysis, abductive unification, and nonmonotonic equality. *Artificial Intelligence, 34*, 275–295.

Conant, J. B. (1957). The overthrow of the phlogiston theory: The chemical revolution of 1775-1789. In J. B. Conant, L. K. Nash, D. Roller, & D. H. D. Roller (Eds.), *Harvard case histories in experimental science.* Cambridge, MA: Harvard University Press.

Conant, J. B., Nash, L. K., Roller, D., & Roller, D. H. D. (Eds.). (1957). *Harvard case histories in experimental science.* Cambridge, MA: Harvard University Press.

Forbus, K. D. (1984). Qualitative process theory. *Artificial Intelligence, 24*, 85–168.

Forbus, K. D., & Gentner, D. (1986). Learning physical domains: Toward a theoretical framework. In Michalski, R. S., Carbonell, J. G., & Mitchell, T. M (Eds.), *Machine Learning: An Artificial Intelligence Approach, Volume II.* Los Altos, CA: Morgan Kaufmann.

Guerlac, H. (1961). *Lavoisier — the crucial year — the background and origin of his first experiments on combustion in 1772.* Ithaca, NY: Cornell University Press.

Ihde, A. J. (1980). Priestley and Lavoisier. In *Joseph Priestly symposium, Wilkes-Barre, Pa., 1974.* London: Associated University Presses, Inc.

Josephson, J. R., Chandrasekaran, B., Smith,Jr., J. W., & Tanner, M. C. (in press). A mechanism for forming composite explanatory hypotheses. *IEEE Transactions on Systems, Man and Cybernetics, Special Issue on Causal and Strategic Aspects of Diagnostic Reasoning.*

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

O'Rorke, P. (1987). LT revisited: Experimental results of applying explanation-based learning to the logic of Principia Mathematica. In *Proceedings of the Fourth International Workshop on Machine Learning*, (pp. 148–159). Irvine, CA: Morgan Kaufmann Publishers, Inc.

O'Rorke, P., & Cain, T. (1988). Explanations involving emotions. to appear in *Proceedings of the AAAI-88 Workshop on Plan Recognition*, St. Paul, MN: Morgan Kaufmann Publishers, Inc.

Peirce, C. S. S. (1931-1958), *Collected Papers of Charles S. Peirce (1839-1914).* Hartchorne, C., Weiss, P., & Burks, A. (Eds.)., Cambridge, MA: Harvard University Press.

Pople, H. E. (1973). On the mechanization of abductive logic. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, (pp. 147–152).

Reggia, J. A., & Nau, D. S. (1984). An abductive non-monotonic logic. In *Proceedings of the Workshop on Non-Monotonic Reasoning*, (pp. 385–395).

Rose, D., & Langley, P. (1986). Chemical discovery as belief revision. *Machine Learning, 1*, 423–452.

Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively.* Lawrence Erlbaum and Associates.

Thagard, P. (in press-a). The conceptual structure of the chemical revolution. *Philosophy of Science.*

Thagard, P. (in press-b). Explanatory coherence. *The Behavioral and Brain Sciences.*

Zytkow, J. M., & Simon, H. A. (1986). A theory of historical discovery: The construction of componential models. *Machine Learning, 1*, 107–137.