Theory Formation by Abduction: A Case Study Based on the Chemical Revolution

Paul O'Rorke (ororke@ics.uci.edu)
Steven Morris (morris@ics.uci.edu)
David Schulenburg (schulenb@ics.uci.edu)
Department of Information and Computer Science
University of California, Irvine
Irvine, CA 92717

(Technical Report Number 89-25)
August 9, 1989

Last Revised: March, 1990

Theory Formation by Abduction: A Case Study Based on the Chemical Revolution

Paul O'Rorke (ororke@ics.uci.edu)
Steven Morris (morris@ics.uci.edu)
David Schulenburg (schulenb@ics.uci.edu)
Department of Information and Computer Science
University of California, Irvine
Irvine, CA 92717

Abstract

Abduction is the process of constructing explanations. This chapter suggests that automated abduction is a key to advancing beyond the "routine theory revision" methods developed in early AI research towards automated reasoning systems capable of "world model revision" — dramatic changes in systems of beliefs such as occur in children's cognitive development and in scientific revolutions. The chapter describes a general approach to automating theory revision based upon computational methods for *theory formation by abduction.* The approach is based on the idea that, when an anomaly 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 chapter shows how some of Lavoisier's key insights during the Chemical Revolution can be viewed as examples of theory formation by abduction.

Contents

List of Figures

List of Tables

1 World Model Revision

Much of the recent progress in AI has been characterized by the slogan *in the knowledge lies the power* (Feigenbaum, 1979). The performance of AI systems has become more impressive as the systems have become increasingly knowledge intensive. Unfortunately knowledge-intensive systems are bound to exhibit imperfect behavior when they are based upon imperfect knowledge. Existing AI systems often exhibit fragility or brittleness as a result. For this reason, revision methods are needed to correct and extend knowledge that is incorrect or incomplete.[1]

"Routine belief revision" methods already exist as a result of progress in AI research. Automated reasoning systems have used methods such as contradiction backtracing (Shapiro, 1981), dependency-directed backtracking, and truth maintenance (Doyle, 1979) in a large number of tasks requiring simple changes in systems of beliefs.

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. This sort of "world model revision" is at the more difficult, more creative end of the spectrum of belief revision problems.[2] 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.

Until recently, it has not been clear how to advance toward reasoning systems capable of world model revision. The claim of this chapter is that computational methods for *theory formation by abduction* can provide a basis for world model revision. Abduction is the process of constructing explanations (Charniak, 1988; Charniak & McDermott, 1986; Josephson, Chandrasekaran, Smith Jr. & Tanner, 1987; Peirce, 1931-1958; Pople, 1973; Reggia, Nau & Wang, 1983; Schank, 1986). This chapter focuses on a theory-driven form of abduction that can be used to derive explanations of anomalous observations, given rules and facts encoding a general theory and the details of a specific situation. If a prediction of a given theory contradicts an observation, the approach to revision advocated here involves explaining the observation in terms of basic principles. We claim that the process of finding an explanation can lead by abductive inference to new hypotheses that can form crucial parts of new theories.

To support this claim, we give a detailed description of a computer simulation viewing one of Lavoisier's key insights in the chemical revolution as an example of theory formation by abduction. In Section 2 we present some background on this particular world model revision. Then in Section 3 we show how advances in qualitative physics provide a language

[1] In this paper, *knowledge base*, *model*, *theory*, and *belief system* are used as roughly interchangeable terms since knowledge bases can be viewed as models, theories, or sets of beliefs.

[2] McDermott and Doyle (1980) first made the distinction between routine belief revision and world model reorganization. They also present a modal approach to the formalization of non-monotonic reasoning and routine belief revision.

for representing some aspects of chemical processes in the form of rules and facts. In Section 4 we discuss our overall framework for theory revision. In particular, we discuss how abduction can provide a means for theory formation. In Section 5, we illustrate how our abduction method generates qualitative chemical explanations. These sections then put us in a position to show in Section 6 how an observation in conflict with the phlogiston theory can lead via abduction to hypotheses that correspond to a key insight that contributed to Lavoisier's revolutionary shift to an oxygen theory of combustion. Section 7 discusses related work and Section 8 provides a brief summary and conclusion.

2 The Chemical Revolution

James Bryant Conant argues in his introduction to the *Harvard case histories in experimental science* (Conant, Nash, Roller & Roller, 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 (e.g., see Langley, Simon, Bradshaw, and Żytkow, 1987; Thagard, 1988). 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, but to improve our understanding of how theory shifts occur.

In this chapter, we present some initial results of a case study of the chemical revolution — the replacement of the phlogiston theory by the oxygen theory (1775 - 1789). This particular theory shift has attracted a great deal of interest partly because it occurred in the early days of chemistry, while the theories and experiments were still close to common knowledge and everyday experience. In addition, a great deal is known about the chemical revolution because of detailed records left by the scientists involved and thanks to the large number of books and papers on the subject by historians and philosophers of science, e.g., see (Conant, 1957; Guerlac, 1961; Ihde, 1980; Thagard, 1988b).

Prior to the Chemical Revolution, the phlogiston theory of chemistry provided the predominant explanation for 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 flow of phlogiston from the combustible substances into the surrounding air. Calcination is an alchemists' term for the process of changing things to calx or powder by applying heat. It also applies to rusting (now called *oxidation*) and was thought of as a loss of phlogiston from metals.[3] The phlogiston theory thus predicts a decrease in the weight of

[3]Calx is the ashy powder left after a metal or mineral has been calcined. It is the same as rust, a coating or film formed on metals by corrosion or calcination. The reddish rust that appears on iron exposed to air

combusting and calcining substances.

Consequently, Lavoisier, the 18th century French chemist who was the chief protagonist in the chemical revolution, placed great importance on the observation that the weights of some substances increase when they undergo combustion and calcination. Just after this "augmentation" effect was demonstrated conclusively, Lavoisier 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.*[4]

With the help of colleagues such as Joseph Priestly, Lavoisier went on to discover that (contrary to the century-old phlogiston theory) a gas contained in the atmosphere combines with burning combustibles and calcining 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."

3 Representing Qualitative Chemical Knowledge

In this section, we show how advances in research on qualitative physics provide a language for describing some important ideas associated with the chemical revolution. First we present a qualitative process schema for combustion according to the phlogiston theory and then consider how some of the ideas associated with qualitative physics and with phlogiston can be encoded in terms of facts and rules. As described in sections 5 and 6, this encoding enables our abduction method to construct explanations of observations involving changes in the weights of burning and calcinating substances.

3.1 A Qualitative Process Description of Combustion

Qualitative process (QP) theory (Forbus, 1984) provides a language for describing qualitative changes due to processes acting on quantities. Table 1 shows a QP representation of a fragment of Stahl's phlogiston theory. This representation is intended to capture the phlogiston theorist's notion that combustion is similar to a "flow" of phlogiston from a combustible substance to the surrounding air.

---

and moisture is the most familiar example.

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

Table 1: A qualitative process description of the phlogiston theory of combustion.

Process: combustion

Individuals:
- combustible a complex-stuff
- phlogiston a simple-stuff
- air a gas

Preconditions:
- ∼wet(combustible)
- component-of(phlogiston, combustible)
- surrounds(air, combustible)

QuantityConditions:
- A[temp(combustible)] $\geq$ A[flashpoint(combustible)]
- A[amount-of-in(phlogiston, combustible)] $>$ ZERO
- A[amount-of-in(phlogiston, air)] $<$ A[capacity-of-for(air, phlogiston)]

Relations:
- Let combustion-rate be a quantity
- combustion-rate $\propto_+$ (A[capacity(air, phlogiston)] $-$ A[amount-of-in(phlogiston, air)])
- combustion-rate $\propto_+$ amount-of-in(phlogiston, combustible)

Influences:
- I $-$ (amount-of-in(phlogiston, combustible), A[combustion-rate])
- I $+$ (amount-of-in(phlogiston, air), A[combustion-rate])

Individual-view: complex-stuff

Individuals:
- complex a substance
- $\{S_i \mid S_i$ a substance$\}$

Preconditions:
- components($\{S_i\}$, complex)

QuantityConditions:
- $\forall$ $S_j \in \{S_i\}$, A[amount-of-in(S_j, complex)] $>$ ZERO

Relations:
- There is p $\in$ piece-of-stuff
- made-of(p, complex)
- $\Sigma_{\{S_i\}}$ amount-of-in(S_i, p) $=$ amount(p)

This framework represents processes as frames or schemata called *qualitative process descriptions.* These frames contain knowledge about the objects (*individuals*) involved in the process, as well as specific knowledge about the process itself. Objects are also described using additional frames known as *individual views.*

The **individuals** slot of a QP schema specifies the objects associated with a process or individual view. In the qualitative description of *combustion*, the individuals include a piece of some combustible substance (e.g., a chunk of charcoal), a piece of phlogiston, and a volume of air. In the individual view of *complex-stuff*, the relevant individuals include some complex substance and a set of component substances.

Among other things, the **preconditions** slot of the combustion schema captures the phlogiston theorist's belief that only complex substances can burn. According to phlogiston theory, all combustibles are compounds containing the element *phlogiston.* The **quantity-conditions** state that in order for combustion to occur, the combustible must be "hot enough", there must be some phlogiston in the combustible, and the surrounding air must not be "saturated" by phlogiston. The **relations** state that the combustion rate is qualitatively proportional to the remaining capacity of the air for phlogiston and to the phlogiston content of the combustible substance. The **influences** state that phlogiston is leaving the combustible and "escaping" into the air, and that this flow is directly influenced by the *combustion-rate.*

The individual view of *complex-stuff* states that there is a set of substances, namely the components of the *complex-stuff,* and that the sum of the amount of each of these substances equals the amount of the piece of *complex-stuff.*

3.2 Aspects of the Phlogiston Theory

The qualitative process description of the phlogiston theory of combustion sketched in the previous section is intended to capture a number of inferences and explanations made by phlogiston theorists. In this section, we simplify the theory and identify a fragment relevant to predicting and explaining why the weight of charcoal decreases as it burns.

Table 2 presents two classes of qualitative laws that capture important aspects of the phlogiston theory. The first class (GL4, GL6, GL7, and GL8) is concerned with certain basic properties of substances. Rule GL4 states that the weight of any substance is proportional to the amount of the substance. Rules GL6, GL7, and GL8 state that the amount of a complex substance is equal to the sum of the amounts of its components.

The other class of laws captures certain aspects of the phlogiston theorist's view on the nature of combustion and calcination. In this view, all combustible substances are complex substances that contain phlogiston. In our qualitative process description of combustion, rule GL5a states that combustion is a process that negatively influences the amount of phlogiston in charcoal. That is, if combustion is active it drives down the amount of phlogiston in a partially burned piece of charcoal. Similarly, Rule GL5b states

Table 2: Some key laws of the phlogiston theory.

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).

Note: 'm-c' abbreviates 'mercurius calcinatus' and stands for a piece of partially calcinated red calx of mercury.

---

that calcination drives down the amount of phlogiston in a partially calcinated piece of mercury. The laws in Table 2 encode a fragment of the knowledge that a phlogiston chemist may have used in reasoning qualitatively about chemical phenomena.

3.3 General Laws of Qualitative Physics

This section presents some general laws of qualitative process theory (see Table 3). These laws are important in common-sense reasoning about the physical world. *The law of direct influences* (GL1) 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."

The *laws of indirect influences* (GL2a and GL2b) are meant to capture the notion that a quantity may change because it is qualitatively proportional to some other quantity. A qualitative proportionality may be either positive or negative. If there is a positive qualitative proportionality, a change in one quantity may be accounted for by a similar change in some other quantity. In the case of a negative qualitative proportionality, a change in one quantity may be accounted for by an opposite change in another quantity.

The *law of sums* (GL3) 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

Table 3: Some general laws of qualitative physics encoded as rules.

Direct Influences:

GL1: deriv-sign(Q1, Sign) $\leftarrow$ process(Process), active(Process), influence(Process, Q1, Sign).

Indirect Influences:

GL2a: deriv-sign(Q1, Sign) $\leftarrow$ qprop(Q1, Q2, pos), deriv-sign(Q2, Sign).

GL2b: deriv-sign(Q1, Sign1) $\leftarrow$ qprop(Q1, Q2, neg), deriv-sign(Q2, Sign2), opposite(Sign1, Sign2).

The Law of Sums:

GL3: qprop(Q, Q_i, pos) $\leftarrow$ qty-eq(Q, qty-sum(Qs)), member(Q_i, Qs).

Note:

In GL1, "deriv-sign(Q1) = Sign" means "the sign of the time derivative of quantity Q1 is Sign."
In GL2a, "qprop(Q1, Q2, pos)" means "quantity Q1 is positively qualitatively proportional to quantity Q2."
In GL3, "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. Also, "member(Qi, Qs)" means "Qi is a member of the list of Qs."

---

which is the second quantity. For example, qprop(weight(body), Q_i, pos) $\leftarrow$ qty-eq(weight(body), qty-sum([weight(lean-body-mass),weight(other-body-mass)])), member(weight(lean-body-mass), [weight(lean-body-mass),weight(other-body-mass)]).

Note that the "implications" in these "laws" are somewhat ambiguous. The implication in the "law of sums" should be interpreted as material implication, whereas the implications in the "laws of influences" should be interpreted as specifying potential causal associations. Used in backward chaining, these rules specify possible causes for events. Used in forward chaining, they predict potential consequences.

Inferences based on the "laws of influences" typically focus on one aspect of a situation under the assumption that other aspects can be safely ignored. In particular, inferences involving a change in some quantity ignore other potential influences or proportionalities involving the affected quantity. In the case of the "laws of indirect influences," a quantity may be qualitatively proportional to another quantity, and this second quantity may be changing, but this change does not necessarily completely determine what will happen to the first quantity. In the case of the "law of direct influences," an active process may be driving a quantity up or down, but that does not rule out the possibility that there are other direct or indirect influences acting in the opposite direction.

A classic example of a set of conflicting influences involves a bathtub with a faucet valve open but with the drain open as well. The water level in the tub is driven upward by the water flowing in through the faucet but it is simultaneously driven downward by the water flowing out through the drain. Forward chaining on rules like the ones in the table could be used to predict possible consequences, such as that the water level in the tub may go up (or down). Alternatively, backward chaining on these rules could be used to generate possible causes, e.g.: explaining why the level of the water is observed to increase (or decrease).

4 Abduction, Hypothesis Formation, and Theory Revision

The previous section described the kinds of knowledge needed to explain certain observations according to the phlogiston theory. In this section we describe how this sort of knowledge can be used to construct explanations by a process of abductive inference. Next, we explain how this form of abduction can be used to generate hypotheses and we describe a method that uses this capability in theory revision.

According to Peirce (1931-1958), abduction is explanatory hypothesis generation. Peirce's formulation of abduction was basically: "The surprising fact, C is observed; but if A were true, C would be a matter of course, hence there is reason to suspect that A is true." An analysis of Peirce's views on abduction and hypothesis formation and evaluation may be found in Thagard (1981). Interestingly, while Peirce originally intended the term abduction to apply only to the initial formulation of hypotheses, AI researchers usually include evaluation and acceptance as part of abduction. We conform to this convention and use the term abduction loosely as shorthand for methods for constructing and evaluating explanations.

4.1 Theory-driven Abduction

AI researchers have cited Peirce's notion of abduction as the basis for a number of different methods and systems. In this section we describe the particular form of abduction ("theory-driven abduction") used in our case study of the chemical revolution. This approach to abduction is related to the philosophical view of explanations as deductive arguments in which the thing to be explained follows from a set of general laws and specific facts. Hempel (1965) calls explanatory accounts of this kind "explanations by deductive subsumption under general laws or *deductive-nomological explanations.* (The root of the term nomological is the Greek word *nomos* for law.)" The form of abduction explored here could be considered to be a deductive-nomological form of abduction, viewing explanations as deductive proofs. The proofs show how observations follow from sets of rules and facts that encode general theories and facts that describe specific situations.

The abduction machinery used here is closely related to theorem provers, and, in particular, to the standard technique called *backward chaining* see, e.g., (Charniak, Riesbeck, McDermott & Meehan, 1987). In this technique, a query *C?* is used to generate a query *A?* by backward chaining on a rule $A \rightarrow C$. In technical terms, this is done by first "unifying" the query *C?* with the conclusion C of the rule. Unification produces a substitution θ that shows how to bind variables in C and C? so as to make them identical. This substitution is then applied to the rule's antecedent A. The result, $A\theta$, is taken as a new query $A?$. This query may "ground out" by unifying with known facts (statements "known to be true" and given as input to the abduction engine), or it may lead to new queries by way of additional backward chaining. In using backward chaining for abduction, observations to be explained are viewed as queries, and general theories and other observations are expressed in terms of rules and facts. Backward chaining attempts to reduce the observations to known facts by way of the rules contained in the theory.

4.2 Abduction and Hypothesis Formation

In order to see how this form of abduction may be used for hypothesis formation, it is important to distinguish between the process that constructs explanations and the resulting explanations. Explanations may be deductive even when the process of constructing them is not deductive. A conclusion may follow deductively from a set of assumptions given the truth of those assumptions but the process of generating the assumptions required to complete the proof may be non-deductive.

In our particular abduction engine, the process of backward chaining on an observation produces partial proof trees. The leaves of these trees may or may not correspond to known facts. In some cases, backward chaining "grounds out" so that *all* of the leaves of a proof tree unify with facts given as part of the input to the abduction engine. For example, Section 5 presents a proof that explains the decrease in a piece of charcoal's weight as it burns. This explanation is derived by backward chaining on the rules given in Section 3 that describe the phlogiston theory, such as the rule that combustion drives the amount of phlogiston in a piece of charcoal down. The proof "grounds out" in statements from this theory and in statements that encode observations (e.g., that the charcoal is burning).

However, when used in constructing explanations, backward chaining often fails to produce complete proof trees. In this case, the ungrounded leaves of the partial proof trees correspond to the explanatory hypotheses generated in Peirce's formulation of abduction. If the propositions corresponding to these leaves were true, the observation would follow, and so there is some reason to suspect that they are true. Yet, even if no better explanation of the observation can be found, the leap to this conclusion is a non-deductive, *abductive inference.* Section 6 shows how this sort of abductive hypothesis formation can be used to generate aspects of the oxygen theory by explaining "augmentation effects," such as the observation that a metallic calx gains weight in calcination.

4.3 Revising Theories Using Abductive Hypothesis Formation

Now we can sketch our approach to theory revision through abductive hypothesis formation. The need for revision is typically recognized when a theory is found to contradict new observations. The task is then to determine what revisions will give a new theory that is in accord with observation. Most approaches to theory revision involve direct transformations that produce the new theory from the original "old" theory. These tranformations are generally very much like "editing" or "tweaking." Two combinatorial problems occur in these transformations. One involves the identification of the erroneous subset of the original theory, and the other involves the identification of the correct changes in the erroneous parts of the original theory. In some situations, these combinatorics are likely to overwhelm editing approaches to theory revision and there is some evidence that people employ a different approach. In Shrager and Klahr's "instructionless learning" experiments, subjects were asked to "figure out" complex programmable devices. Shrager (1987) comments:

> *... we observed that between interactions with the BigTrak, subjects changed their theory of the device. A number of empirical generalizations seem to hold about the nature of these changes... Instead of trying to determine in detail what led to a failed prediction, subjects usually observed what (positive behavior) took place and changed their theory according to that observation...*

We believe that the approach to theory revision explored in the present paper is compatible with Shrager's results. When a surprising observation contradicts a prediction of the original theory, our approach involves retracting questionable beliefs. However, one need not start by trying to identify an individual incorrect belief or even a small set of culprits. Instead, we assume that the initial theory has some internal structure and that more general fundamental principles can be separated from relatively specific, less basic statements. A "core" subset of the original theory, a set of basic statements having nothing to do with the anomaly, is retained while less central beliefs are suspended. Our approach involves explaining the unexpected new observation in terms of the remaining, relatively solid basic principles. As we will see in Section 6, this explanation process can generate hypotheses, suggesting extensions to the basic principles. If these hypotheses are added to the basic principles, the resulting set of rules and facts is a candidate revised version of the original theory.

This approach to theory revision is sketched in Figure 1 using Venn diagrams. In the first stage (a) of theory revision an anomaly is noted. A new observation contradicts a prediction of the old theory, as indicated by the X linking a point in the old theory and a point outside of it. In the next stage (b) the old theory is reduced to the core subset.[5] Starting from this subset, an explanation of the new observation is abduced with hypotheses being introduced in the process. These hypotheses then form the basis for

[5]Notice that neither the prediction nor the surprising observation are included in the reduced core subset of the original theory. The circles and ellipses designate theories closed under deductive inference. The figure captures the notion that neither the prediction nor the contradictory observation should be implications of the core theory.

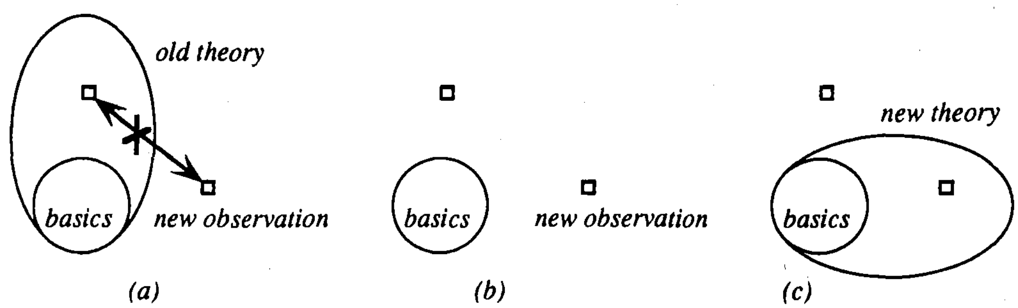

Figure 1: Theory revision using abduction for hypothesis formation

extensions to the core theory resulting in a new theory (c). This revised theory no longer makes the erroneous prediction of the old theory.

We will not explore the initial step of falling back on basic principles and shrinking the original theory in this chapter. Nor do we explore the evaluation of the original theory and the candidate revision. Instead, we focus on the step from Figure 1(b) to Figure 1(c). We concentrate on the claim that the process of explaining unexpected new phenomena can lead by abductive inference to new hypotheses which can form crucial parts of new theories. We substantiate this claim by showing how an aspect of the oxygen theory of combustion and calcination can result from revising the phlogiston theory using this framework.

5 The Phlogiston Account of Burning Charcoal

Now we are in a position to illustrate the use of the qualitative physical laws and the phlogiston theory in the construction of explanations, showing how our "abduction engine" generates explanations of an observation by attempting to reduce it to known facts using general laws. We have implemented this approach in AbE, a PROLOG meta-interpreter that uses best-first heuristic search to construct explanation trees and evaluate partial explanations. AbE's heuristic evaluation function is based upon the "weighted abduction" method proposed by Stickel in (Hobbs, Stickel, Martin & Edwards, 1988). This section shows how the system constructs an explanation of the drop in the weight of burning charcoal, in accord with the phlogiston theory.

AbE is given the observation that, upon burning, the weight of some charcoal decreases. This is expressed as a statement (labeled O1 in Table 4) that the sign of the derivative of the weight of the charcoal is negative. The system is also given some specific facts (CF1-5) that combustion is occurring and that charcoal is a complex substance containing phlogiston and ash. (This was the model of charcoal held by the phlogiston chemists.) In addition, the system is given the general laws of qualitative physics and the phlogiston theory described earlier in Tables 2 and 3. AbE is asked to explain the observation by

Table 4: The weight of charcoal decreases as it burns.

Observation:		
	O1:	deriv-sign(weight(charcoal), neg).
Case facts:		
	CF1:	process(combustion).
	CF2:	active(combustion).
	CF3:	complex(charcoal).
	CF4:	component(phlogiston, charcoal).
	CF5:	component(ash, charcoal).

using the given laws to connect the observation to the given facts. The output of the system is a set of proof trees like that shown in Figure [ExplTree1]. In the process of constructing explanations, the observation is treated as a query that gives rise to new queries by backward chaining on rules representing logical and causal laws. The remainder of this section traces the construction of the tree, visiting the nodes in the order shown by the labels in Figure 2.

The initial query addressed by AbE is *why is the weight of the charcoal decreasing?* According to the laws of indirect influences (GL2), a change in some quantity may be explained by a change in some other quantity provided those quantities are qualitatively proportional. This raises the question of whether the decrease in the weight of the charcoal may be explained in terms of a decrease in some amount that is positively qualitatively proportional to the weight of the charcoal. 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).

At this point, the question is *why is the amount of charcoal decreasing?* To explain this change, the system again attempts to find a positive qualitative proportionality between the amount of charcoal and some other decreasing quantity (using GL2). An appropriate proportionality is found using the law of sums (GL3). This states that some quantity Q is proportional to some other quantity Q_i if Q is equal to the sum of some set of quantities Qs and Q_i is a member of that set. In this case, Q is the amount of charcoal.

The question is now *whether there is some set of quantities whose sum is equal to the amount of charcoal.* AbE answers this in terms of its knowledge about complex substances. In particular, the system knows that the overall amount of a complex substance is equal to the sum of the amounts of the components of the substance (GL6). In Figure 2, the amount of charcoal is shown as a quantity sum over a list consisting of the amount of phlogiston in the charcoal and the amount of ash in the charcoal. This list of components is actually derived through several applications of rules GL7 and GL8 (which, if shown in Figure 2,

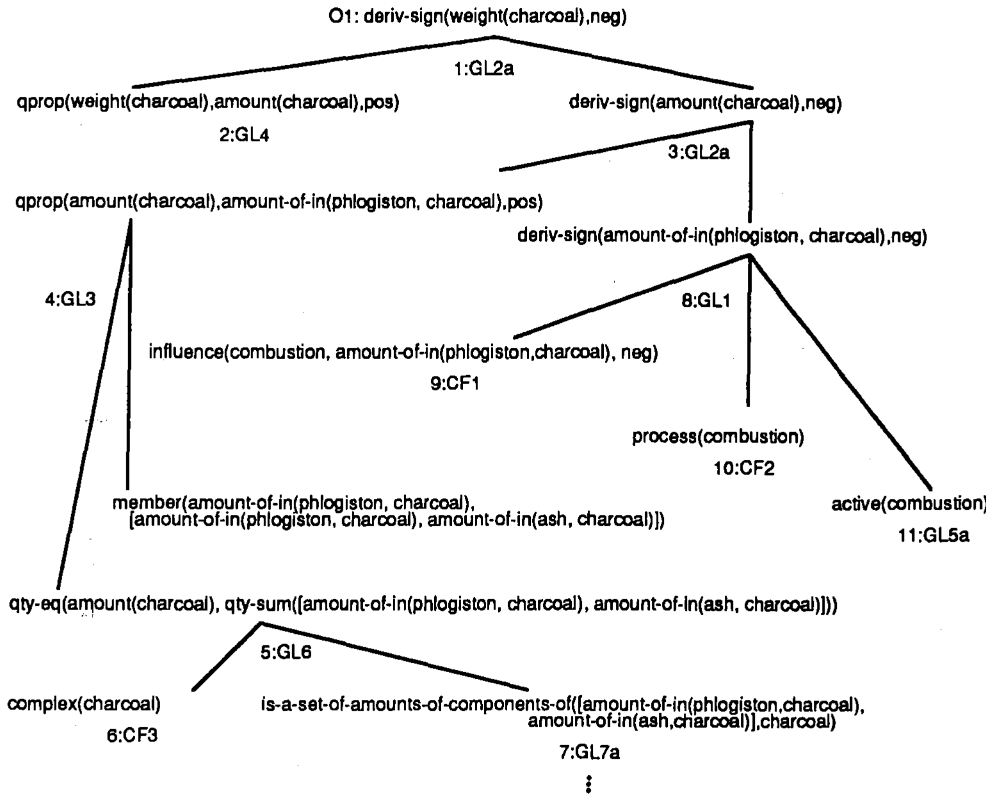

Figure 2: An explanation of the weight decrease in burning charcoal.

would construct a subtree beginning at node 7). In particular, the two applications of rule GL8 ground out using the case facts CF3 (charcoal is a complex substance), CF4 (phlogiston is a component of charcoal), and CF5 (ash is a component of charcoal). The result is a proof tree for node 4, which states that the overall amount of the charcoal is positively qualitatively proportional to the amount of phlogiston in the charcoal.

The question now is *whether the amount of phlogiston in the charcoal is decreasing.* The law of direct influences (GL1) can be used to explain this decrease, assuming that an active process can be found to have a negative influence on the amount of phlogiston in the charcoal. At this point, the system completes its explanation by recognizing that, according to facts of the case and a key statement in the phlogiston theory, combustion is an active process that negatively influences the amount of phlogiston in the charcoal.

AbE also generates a similar proof, in which the amount of charcoal is seen as proportional to the amount of ash. However, it cannot ground this proof at node 11 via a direct, negative influence of combustion on the amount of ash, because this influence is not a fact. Consequently, in the course of its heuristic search, AbE's evaluation function ranks this incomplete explanation less favorably than the explanation presented in Figure 2.

6 Abduction of Aspects of the Oxygen Theory

In Section 3, we saw that recent progress in knowledge representation and automated reasoning makes it possible to capture key ideas contained in early chemical theories such as the phlogiston theory. In Section 4, we claimed that recent progress on automated abduction makes it possible to capture significant aspects of the reasoning that occurred in the chemical revolution. In Section 5, we showed how abduction on a simplified phlogiston theory can generate an explanation that seems to capture the phlogiston theorists' views that substances lose weight when they burn or calcine because they lose phlogiston. This explanation was constructed by reducing the observation (that a piece of charcoal lost weight) to "known facts" (given to AbE *before* it started searching for an explanation).[6] However, in Section 4 we claimed that abduction could also be used to go beyond deductive inference to form the kinds of hypotheses involved in major theory shifts.

In this section, we describe how abduction can generate a crucial aspect of the oxygen theory. Let us assume as *given* the phlogiston theory account of combustion and calcination, along with the observation and case facts shown in Table 5. Ignore for the moment that several items in Table 5 are crossed out. Recall that m-c is an abbreviation for mercurius calcinatus, which is partially calcinated 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," calcined out of metals.

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 and the case facts responsible for the contradiction have been identified and deleted as indicated by the offending statements crossed out in the table.[7]

Assume, then, that our abduction engine AbE is given this reduced phlogiston theory along with the observation and case facts shown in Table 5. The altered theory and observation make no mention of phlogiston, so that it is no longer considered as an essential component of combustible substances or as involved in combustion or calcination.

[6]PROLOG could have been easily used to construct the explanation in Section 5.

[7]Existing contradiction backtracing (Shapiro, 1981) and dependency-directed backtracking methods (Doyle, 1979) 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 potential culprit should be excised. Basic principles that contribute to many explanations (e.g., conservation laws) should be preferentially retained.

Table 5: Ablation of the phlogiston theory.

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).

Anomaly: The weight of red calx of mercury increases as it calcinates.

Observations:

O1: deriv-sign(weight(m-c), pos).

Case facts:

CF1: process(calcination).

CF2: active(calcination).

AbE is asked to explain, in terms of the given laws of qualitative physics and the ablated phlogiston theory, the observation that the weight of mercurius calcinatus increases (O1) during calcination (CF1 and CF2). The system 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 3 shows one explanation that AbE generates. This explanation is obviously very similar to the explanation of the decrease in the weight of charcoal according to the phlogiston theory discussed earlier. Let us examine how this explanation was constructed.

The initial query is *why is the weight of the mercurius calcinatus increasing?* The system answers this question in terms of an increase in the amount of the substance.

This leads to the question *why is the amount of mercurius calcinatus increasing?* To explain this, AbE tries to find a positive proportionality between the amount of mercurius calcinatus and some other increasing quantity. The system finds an appropriate propor-

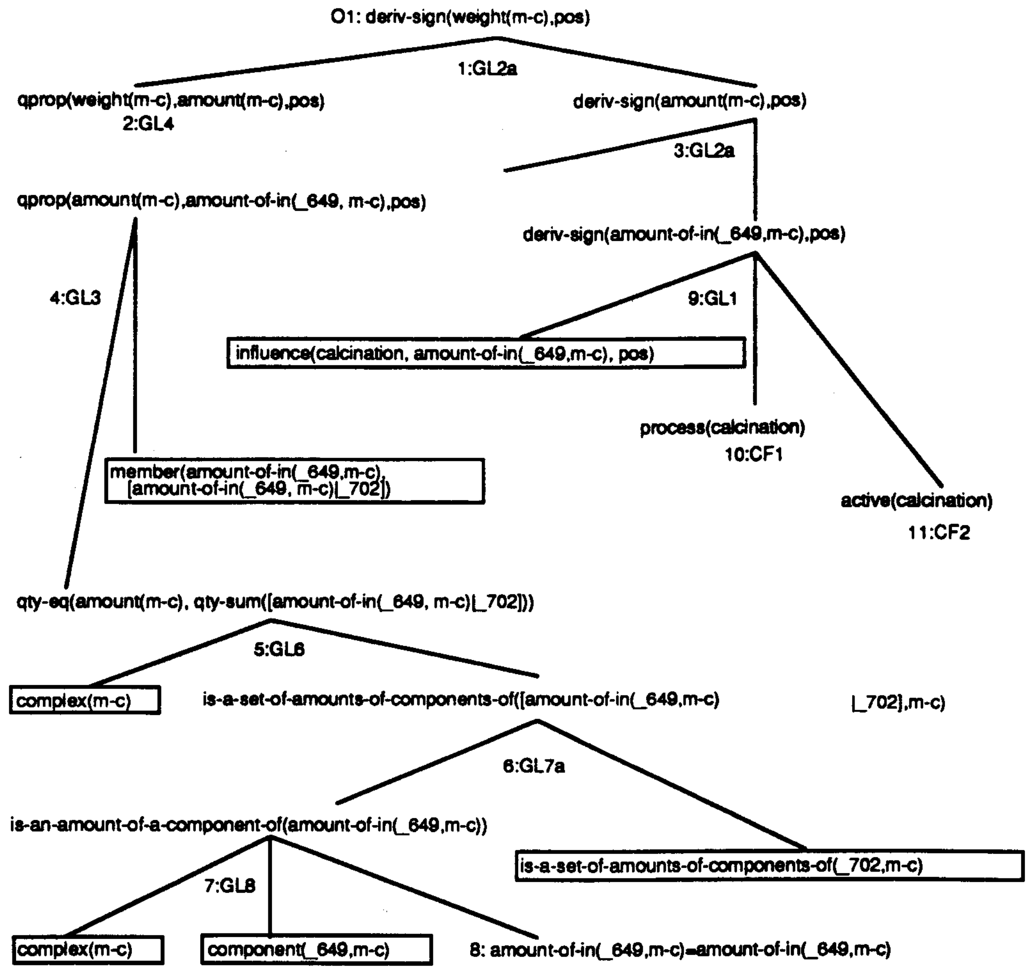

Figure 3: An explanation of the weight increase in mercurius calcinatus during calcination.

tionality using the law of sums (GL3), which it then uses to hypothesize that the amount of mercurius calcinatus is increasing because the amount of one of its components is increasing. Using laws about complex substances (GL6, GL7, GL8), AbE hypothesizes the existence of an unknown component of mercurius calcinatus and the existence of an unknown quantity corresponding to its amount. The system also hypothesizes that there is a set of remaining components and associated amounts, without identifying any particular elements of this set.[8] Unlike the burning charcoal example, in the present case AbE is *not*

[8]These new individuals are represented by Skolem constants generated as a natural consequence of backward chaining. As shown in Figure 3, the set of the component amounts is represented as the list [amount-of-in(_649,m-c) | _702], where _649 is the hypothesized unknown component, and _702 is the remaining subset of component amounts.

given a case fact stating that mercurius calcinatus is complex. Instead, in using the law of sums, AbE hypothesizes that mercurius calcinatus is a complex substance.[9]

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, if one assumes an active process that has 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 has a direct positive influence on the amount of the unknown component.

The construction of the above explanation used only general rules and facts such as the rule that a quantity may have changed as a result of an active process and the fact that the weight of a substance is qualitatively proportional to its "amount". There was no use of knowledge encoding the chemical content of specific qualitative chemical theories, such as the phlogiston or oxygen theory. Instead, AbE employed the basic theory to generate hypotheses corresponding to parts of the oxygen theory by abductive inference. The hypotheses generated in this manner are enclosed in boxes in Figure 3. These abductive inferences correspond to Lavoisier's insight that something was being added during calcination.

Explaining new, surprising observations is a key step in theory revision. In the case of the chemical revolution, Lavoisier's hypothesis that something was added by calcination to the calx of mercury, in conjunction with experimental results of Priestley and others, eventually led him to posit the existence of a hitherto unknown component of air. Lavoisier called this new theoretical entity "oxygen." During the next decade, he and his colleagues worked out a new theory of combustion, calcination, and respiration that eventually displaced the phlogiston theory. This occurred because most chemists of the time were persuaded that the new theory explained the new observations (and re-explained old observations) in a more coherent manner than did modified versions of the phlogiston theory.

7 Relation to Other Work

Our work is related to, and builds on, work on common-sense reasoning about the physical world, qualitative physics, and scientific discovery. The approach fits into the four-stage theoretical framework for learning in physical domains sketched by Forbus and Gentner (1986). The learning taking place in our chemical revolution appears to fit in the third stage ("learning naive physics"). Here we briefly consider its relation to other work in machine discovery.

[9]The apparently tautological node labeled "8:" in Figure 9 reflects the unification of Qi to amount-of-in(_649,m-c) in law GL8.

7.1 STAHL and STAHLP

Recent work in scientific discovery has produced two systems, STAHL (Żytkow & Simon, 1986) and STAHLp (Rose & Langley, 1986), that can automatically detect and correct errors in chemical theories. Both systems represent chemical theories in terms of *reaction and component models*. A reaction is specified by its *input* and *output* substances; i.e., the substances entering the reaction and the substances resulting from the reaction. A component model specifies the components of an individual complex substance as a list of substances.

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. However, in our opinion, such an account of the theory shift would be incomplete if only because the models of the phlogiston and oxygen theories would be 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 their explanations cannot be expressed solely in terms of component models and reactions.

The inputs to STAHL and STAHLp are reactions. From these reactions, they use rules to derive new, *inferred* reactions and component models. An inferred reaction is generated from a parent reaction by two methods: (1) *reduction*: a substance that appears on both sides of the parent reaction is removed on both sides; and (2) *substitution*: a substance in the parent reaction is replaced by the components of that substance as specified by its component model. A component model of a complex substance is inferred in one way only: when an input or inferred reaction has exactly one input substance, the inference is that the outputs of the reaction specify the components of that substance. Both programs, using different techniques, carry out belief revision in an incremental fashion; beliefs are revised when inconsistent sets of reactions or component models are inferred and detected.

We now discuss some aspects of belief revision in STAHL and STAHLp. In STAHL, inconsistent inputs may lead to the inference of component models that result in infinite recursion. For example, the following two component models could be inferred:

```
mercury = calx-of-mercury + phlogiston
 calx-of-mercury = mercury + oxygen
```

Substitution of the second model into the first produces a component model of mercury that is self-referential, and which leads to infinite recursion:

```
mercury = mercury + oxygen + phlogiston
```

STAHL solves this problem by renaming calx-of-mercury in one of the two component models as "calx-of-mercury-proper." This can be seen as a model of the historical practice of chemists casting doubt on the proposed identity of a substance in a reported reaction.

This allows the introduction of new substances, *but only by way of renaming substances already mentioned in reactions.*

In STAHLp, beliefs are revised when the system infers an inconsistent reaction in which either the input or output side has no substances while the other side has one or more substances. The program corrects this situation by revising its input reactions. Each revision involves deleting or adding a substance from one side of a reaction, and an input reaction can have more than one such revision made to it. Belief revision is effected by identifying a set of revisions of the input reactions that satisfies two conditions: (1) a balanced reaction will be inferred from the revised input reactions, and (2) the number of component models that will be changed by the revision is minimal. Once this revision is identified, all beliefs (inferred reactions and component models) that depend on the revised input reactions are deleted; then the revisions to the input reactions are made; and then STAHLp generates the new reactions and component models that follow from the revised theory. The result of the theory revision can be the elimination or modification of previously held component models and inferred reactions, and the addition of new component models and inferred reactions.

Although such revision in STAHLp amounts 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. The system is not capable of hypothesizing the existence of a new substance that has not previously appeared in an input. This is in contrast to AbE, which can hypothesize a new component substance on the basis of general laws concerning qualitative physics, sums of quantities, and complex substances.

7.2 COAST and PHINEAS

Falkenhainer and Rajamoney (1988) describe a closely related approach to theory revision. The PHINEAS system (Falkenhainer, 1988) extends abductive hypothesis formation to include qualitative physical analogies, and the COAST system (Rajamoney, 1988) revises qualitative physical theories involving processes such as evaporation and osmosis, using "explanation-based theory revision" to propose changes in an initial theory in response to an anomaly. The theory revision process takes an "editing" approach, in the sense that it focuses on both the prediction of the initial theory and the surprising observation that contradicts the prediction. Revision rules are used to generate ways of changing the initial theory so that the prediction is no longer made but the unexpected observation is predicted instead. An advantage of this "theory debugging" strategy is that the errors in the initial theory and their corrections are identified together. In related work, Rajamoney (1989) describes a method for using "exemplars" to guide theory revision, in which he uses qualitative process schemata for a phlogiston theory in an example. However, the revisions proposed by his method are essentially "patches" of the phlogiston theory, and the existence of a new substance (oxygen) is not hypothesized.

7.3 ECHO and PI

Thagard (1989) presents a theory of explanatory coherence and a connectionist implementation. His ECHO program 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. With regard to the question of where such theories might have come from in the first place, Thagard hints that PI (Holland, Holyoak, Nisbett & Thagard, 1986) might be able to construct them. In another paper (Thagard, 1988b), he examines 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. He also suggests that the mechanisms for concept formation and rule abduction present in PI can be used to form conceptual networks that can chart the conceptual changes which occurred during the chemical revolution. Our work on AbE has moved beyond this, showing a detailed example of how abduction, in concert with ideas from qualitative physics, can make some crucial inferences associated with the discovery of oxygen.

8 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 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 chapter 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 advancing beyond "routine theory revision" towards automated reasoning systems capable of "world model revision."

Acknowledgements

This is a revised and expanded version of an earlier paper (O'Rorke, Morris & Schulenburg, 1989). The ideas in this chapter have evolved in discussions with many members of the machine learning community at the University of California, Irvine. Special thanks are due to Pat Langley, Deepak Kulkarni, and Don Rose for numerous discussions of scientific discovery which sparked our interest in modeling the chemical revolution. Discussions with Paul Thagard on Peirce, abduction, the chemical revolution, and on scientific revolutions in general were also inspirational. Thanks to Pat Langley and Jeff Shrager for constructive criticism of an earlier draft. This chapter is based on work supported in part 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.

References

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

Charniak, E., & McDermott, D. V. (1986). *Introduction to Artificial Intelligence.* Reading, MA: Addison-Wesley.

Charniak, E., Riesbeck, C. K., McDermott, D. V., & Meehan, J. R. (1987). *Artificial intelligence programming* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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* (pp. 65-116). Cambridge, MA: Harvard University Press.

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

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence, 12*, 231–272.

Falkenhainer, B. (1988). *Learning from physical analogies: A study in analogy and the explanation process.* (Ph.D. thesis, Technical report number UIUCDCS-R-88-1479). Urbana-Champaign, IL: University of Illinois, Department of Computer Science.

Falkenhainer, B., & Rajamoney, S. A. (1988). The interdependences of theory formation, revision, and experimentation. *Proceedings of the Fifth International Conference on Machine Learning* (pp. 353–366). Ann Arbor, MI: Morgan Kaufmann.

Feigenbaum, E. A. (1979). Themes and case studies of knowledge engineering. In D. Michie (Eds.), *Expert systems in the micro electronic age* (pp. 3–25). Edinburgh: Edinburgh 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 R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 311–348). San Mateo, CA: Morgan Kaufmann.

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

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Eds.), *Aspects of scientific explanation, and other essays in the philosophy of science* (pp. 331–496). New York: MacMillan.

Hobbs, J. R., Stickel, M., Martin, P., & Edwards, D. (1988). Interpretation as abduction. *Proceedings of the Twenty Sixth Annual Meeting of the Association for Computational Linguistics* (pp. 95–103). Buffalo, NY: The Association for Computational Linguistics.

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

Ihde, A. J. (1980). Priestley and Lavoisier. *Joseph Priestly Symposium, Wilkes-Barre, Pa., 1974* (pp. 62–91). London, England: Associated University Presses.

Josephson, J. R., Chandrasekaran, B., Smith Jr., J. W., & Tanner, M. C. (1987). A mechanism for forming composite explanatory hypotheses. *IEEE Transactions on Systems, Man and Cybernetics, Special Issue on Causal and Strategic Aspects of Diagnostic Reasoning, 17*(3), 445–454.

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

McDermott, D. V., & Doyle, J. (1980). Non-monotonic logic I. *Artificial Intelligence, 13,* 41–72.

O'Rorke, P., Morris, S., & Schulenburg, D. (1989). Abduction and world model revision. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 789–796). Ann Arbor, MI: Lawrence Erlbaum and Associates.

Peirce, C. S. S. (1931-1958). *Collected papers of Charles Sanders Peirce (1839-1914).* Cambridge, MA: Harvard University Press.

Pople, H. E. (1973). On the mechanization of abductive logic. *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 147–152). Stanford, CA: Morgan Kaufmann.

Rajamoney, S. (1988). *Explanation-based theory revision: An approach to the problems of incomplete and incorrect theories.* (Ph.D. thesis, Technical report number UILU-ENG-88-2264). Urbana-Champaign: University of Illinois, Coordinated Science Laboratory.

Rajamoney, S. A. (1989). Exemplar-based theory rejection: An approach to the experience consistency problem. In A. M. Segre (Ed.), *Proceedings of the Sixth International Workshop on Machine Learning* (pp. 284–289). Ithaca, NY: Morgan Kaufmann.

Reggia, J. A., Nau, D. S., & Wang, P. Y. (1983). Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies, 19*, 437–460.

Rose, D., & Langley, P. (1986). Chemical discovery as belief revision. *Machine Learning, 1*(4), 423–452.

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

Shapiro, E. Y. (1981). *Inductive inference of theories from facts* (Research report number RR192). New Haven, CT: Yale University, Department of Computer Science.

Shrager, J. (1987). Theory change via view application in instructionless learning. *Machine Learning, 2*, 247–276.

Thagard, P. (1981). Peirce on hypothesis and abduction. In K. Ketner (Ed.), *Proceedings of the C. S. Peirce Bicentennial International Congress* (pp. 271–274). Lubbock, TX: Texas Tech University Press.

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

Thagard, P. (1988b). *The conceptual structure of the chemical revolution* (Technical Report). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Thagard, P. (1989). Explanatory coherence. *The Behavioral and Brain Sciences, 12*(3), 435–502.

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